Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
LOWALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Cardiac Conditioning: The Effects and Implications of Controlled and Uncontrolled Respiration: M. R. WESTCOTT AND J. HUTTENLOCHER 353

Response Latencies Produced by Massed and Spaced Learning of a Paired-Associates List: J. BROWN AND M. HUDA 360

Analysis of Double Alternation in Terms of Patterns of Stimuli and Responses: J. W. SCHOONARD AND F. RESTLE 365

Judgments of Surface Illumination and Lightness: J. BECK 368

Mathematical Solution in the Acquisition of a Verbal CR: J. P. DAS 376

The Discrimination of Relative Onset-Time of the Components of Certain Speech and Nonspeech Patterns: A. M. LIBERMAN, K. S. HARRIS, J. A. KINNEY, AND H. LANE 379

The Relation of Apparent Brightness to the Threshold for Differences in Luminance: E. G. HEINEMANN 389

Temporal Aspects of Cutaneous Interaction with Two-Point Electrical Stimulation: E. SCHMID 400

Prediction of Recurrent Sequences as Related to Level of Irrelevant Cues: A. M. BARCH 410

Effects of Various Asymptotic Restrictions on Human Trial-and-Error Learning: R. W. CHAMBERS AND C. E. NOBLE 417

Information Content in Recognition and Recall: R. DAVIS, N. S. SUTHERLAND, AND B. R. JUDD 422

Percentage Estimation of Proportion as a Function of Element Type, Exposure Time, and Task: E. H. SHUFORD 430

Concept Identification as a Function of Language Pretraining and Task Complexity: E. A. RASMUSSEN AND E. J. ARCHER 437

Inferential Behavior in Children: II. The Influence of Order of Presentation: T. S. KENDLER AND H. H. KENDLER 442

American Psychological Association

Vol. 61, No. 5 May 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 61, NO. 5 MAY 1961

CARDIAC CONDITIONING: THE EFFECTS AND IMPLICATIONS OF CONTROLLED AND UNCONTROLLED RESPIRATION [1]

MALCOLM R. WESTCOTT AND JANELLEN HUTTENLOCHER

Vassar College *Bethesda, Maryland*

Previous experimenters have reported the successful conditioning of cardiac responses in human *Ss*, using a neutral CS and electric shock as UCS, but the form of the reported CR has varied. By form is meant both the beat-by-beat or second-by-second activity of the heart during a period of measurement and the overall rate of the activity during that period as compared with the rate during a control period. Conditioned cardiac deceleration based on pairs of beats measured 6 sec. after CS onset has been reported by Notterman, Schoenfeld, and Bersh (1952), while Zeaman, Deane, and Wegner (1954) and Zeaman and Wegner (1954, 1957) have described a CR consisting of an acceleration followed by a more marked depression in rate. In an unpublished study, cardiac conditioning took the form of an acceleration based on the mean rate for the first 5 sec. after CS onset. Each of these findings involves the comparison of the post-CS rate with the pre-CS rate as a reference point. It should be noted that these three different observations of the CR do not necessarily contradict each other, since measurements were made at different times during the CS-UCS interval. They do, however, pose both an empirical problem arising from the absence of clear confirmation or refutation of each other as well as a specific problem for a stimulus- substitution view of conditioning since the typical response to shock is generally considered to be an acceleration. Zeaman and Wegner (1954, 1957) have provided several studies and discussions which explore the drive-reduction view of cardiac conditioning.

In the unpublished study referred

[1] This investigation was carried out in the Laboratory of Social Relations, Harvard University, and the Massachusetts Mental Health Center. It was supported by a research grant (M-1246) from the National Institutes of Health, United States Public Health Service. A portion of this paper was read at the American Psychological Association meetings, September 1957 (Huttenlocher & Westcott, 1957), and a portion was submitted by the senior author as part of a doctoral dissertation, Harvard University, 1958. The authors wish to thank Richard L. Solomon for his counsel throughout these studies.

to above, respiration records were taken concurrently with cardiac recordings, and a study of these records showed that there were conditioned respiratory changes as well as changes in cardiac rate. During the pretests when the CS was presented alone, 5% of the CS presentations were accompanied by sharp intakes of breath; after 10 conditioning trials, an analysis of the first three extinction trials showed 33% of the CS presentations to be accompanied by such sharp inspirations. When those trials accompanied by this respiratory activity were eliminated from the calculations, no conditioned cardiac acceleration was evident.

Thus, it appeared that a clarification of the relationship between respiratory activity and heart rate under neutral conditions was called for. Further, it appeared that if this relationship proved to be a significant one, some control would be necessary in studies of cardiac conditioning.

The aim of this paper is (*a*) to indicate the influence of several different respiratory patterns on cardiac activity, (*b*) to report a cardiac conditioning study in which breathing during the CS-UCS interval was controlled, and (*c*) to show both the form of the conditioned cardiac response and the form of the acquisition curve for such a response when respiration is effectively held constant. The implications of these findings for previously reported results will also be discussed.

Experiment I

A pilot study of the effects of respiration on heart rate under neutral conditions was carried out.

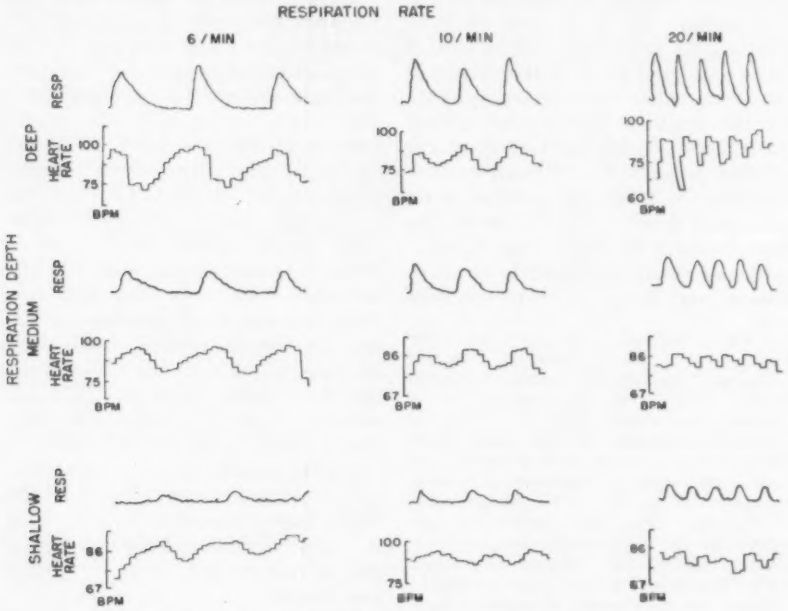


Fig. 1. Concurrent respiration and cardiac recordings for a subject breathing under nine different conditions.

Method

Subjects.—The *S*s were six men free of any history of cardiac difficulty. They ranged in age from 23 to 45 yr.

Apparatus.—Recordings were taken on two channels of a polygraph. Respiration was measured by means of a respirometer around the thorax and heart rate by means of a beat-by-beat cardiotachometer, fed by large electrodes attached to the chest.

Procedure.—The *S* was seated in the experimental room and the recording electrodes attached. Instructions and practice in breathing under nine different conditions were given. The nine conditions were produced by varying three rates of respiration—6/min, 10/min, and 20/min—with three depths of respiration evaluated by *S* as deep, medium, and shallow. After sufficient practice to stabilize the behavior, each of these nine breathing patterns was carried out for 15 cycles by each *S*. In addition, each *S* was instructed to produce isolated gasps of various sorts. Concurrent respiration and cardiac recordings were made throughout the procedure.

Results

The cardiac patterns associated with various breathing patterns showed very little variability among *S*s and almost no variability within an *S* from cycle to cycle. With deep breathing at 6 cycles/min, a change in heart rate on the order of 30 beats per min. (BPM) was observed accompanying each respiration cycle; at 10 cycles/min, dependable changes of 14 BPM, and with 20 cycles/min, orderly fluctuation gave way to a rapid steady rate. Respiration at medium depth produced dependable heart rate changes on the order of 15 BPM, 15 BPM, and 8 BPM for the three speeds, respectively. Shallow breathing yielded heart rate changes dependably associated with breathing at the slow and medium rates only—on the order of 8 and 5 BPM, respectively. For rapid shallow breathing the changes were smaller and irregular. Figure 1 shows an example of each type of breathing and its associated cardiac pattern for a typical *S*.

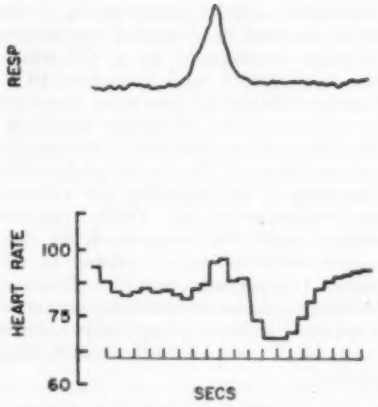


FIG. 2. Concurrent respiration and cardiac recordings of an isolated gasp under neutral conditions.

Isolated gasps also yielded remarkably consistent cardiac patterns within and between *S*s. A single sharp inspiration is shown in Fig. 2 as an example. In the absence of respiratory activity just prior to the gasp, a fairly steady rate of approximately 83 BPM is present. Upon inspiration the rate rises to a peak of 95 BPM, falls to a trough of 70 BPM after 6 sec., and then rises again to the original level.

Discussion

It is clear from these records that there are profound and reliable effects on cardiac rate attributable to respiratory patterns. While the phenomenon of sinus arrhythmia is well known to both physiologists and psychologists, it appears that underestimation of the possible magnitude of fluctuation in cardiac rate arising from respiratory activity, and the probability of the occurrence of such respiratory activity have led to an underestimation of the importance of this phenomenon in studies of cardiac conditioning. Of particular importance is the fact that the sort of

respiratory activity shown in Fig. 2 has been observed as a typical respiratory response conditioned to a CS which has been paired with a noxious UCS. Cardiac activity of the sorts observed in previous studies of cardiac conditioning can readily be identified in the cardiac pattern shown in Fig. 2—an acceleration appearing at the beginning and a larger deceleration at the end. This is a pattern which would be expected if *S*s did, indeed, tend to inspire reliably at CS onset. It is apparent that if cardiac conditioning is to be studied independent of respiratory influences, respiratory control must be a part of such conditioning studies.

Experiment II

An attempt was next made to apply the above findings through a cardiac conditioning study which incorporated rigorous control of respiration with the usual neutral stimulus and shock.

Method

Subjects.—The *S*s were 44 male college students, all free of any history of either cardiac or psychiatric difficulty. They ranged in age from 18 to 24 yr.

Apparatus.—The CS was a house buzzer damped well with cotton and wired through a timer which allowed it to sound for 7 sec. Upon completion of the 7-sec. interval, a relay closed for 0.5 sec. and activated the shock circuit. The shock was delivered from a Reiter Electrostimulator modified to include a Wheatstone Bridge circuit by which changes in *S*'s skin resistance could be monitored, and the voltage altered to maintain a constant shock level. The buzzer and shock could be operated independently. In addition, there was an electric metronome in the experimental room, by which breathing was paced, and a memory drum which exposed bands of color at 3-sec. intervals. The recording apparatus was the same as in Exp. I.

Procedure.—The *S* was seated in the experimental room and the recording and stimulating electrodes attached to his body. The shock was delivered through a circular electrode (1 cm. diameter) on the middle finger of the right hand and a large saline sponge on the right wrist. Neither the metronome nor the memory drum were used in the first part of the experiment. The *S* was told that *E* would deliver a series of electric shocks of 0.5-sec. duration at about 1-min. intervals, beginning very small and each one stronger than the one preceding it. The *S* was to evaluate each shock as to its discomfort on a scale ranging from imperceptible, through uncomfortable, painful, to the limit of tolerance. Each shock was delivered only with *S*'s full knowledge and approval and the series was discontinued when *S* reported his limit of tolerance. The final shock intensity reached usually involved considerable gross bodily reaction on the part of *S* and indicated that he was receiving a genuinely painful shock. The *E*'s own experience with shock in the general range employed supports this inference.

Next the metronome and the memory drum were turned on and *S* was instructed and given practice in the following sequence of activities until he felt comfortable with it and the polygraph recordings showed constant performance from trial to trial: *S* named colors continuously as they appeared on the memory drum at 3-sec. intervals until the buzzer sounded; during the 7-sec. buzzer *S* breathed shallowly at 46 cycles/min, paced by the constantly clicking metronome; at the termination of the buzzer, he returned to color naming.

The *S* was then told that he would be alone for the remainder of the experiment; that he was to begin naming colors as soon as *E* left the room; that from time to time the buzzer would sound and that he was to carry out the breathing activity practiced during the buzzer; that from time to time he would receive an electric shock.

The schedule of trials included 10 presentations of the buzzer alone at random intervals averaging 90 sec., followed by 10 presentations of the buzzer followed by shock. Each shock was at the highest intensity accepted by *S* during the first part of the experiment. On the conditioning trials the 44 *S*s were divided into four groups of 11, each with a different intertrial interval, averaging 1 min., 3 min., 5 min., or 7 min. A Groups × Trials analysis of variance yielded no significant effect of intertrial interval on conditioning so the groups will not be considered separately here.

Measurements of cardiac rate were taken as follows: the mean rate for the 7 sec. immediately preceding CS onset was the prestimulus rate; the mean rate for the 7 sec. of CS presentation was the poststimulus rate; the poststimulus rate minus the prestimulus rate was the response. This analysis yields a net conditioning result. In addition, a

second-by-second analysis was made in order to observe the form of the cardiac activity during the various measurement periods.

Since cardiac rates prior to CS onset varied from *S* to *S* and from trial to trial, a correction was applied to each response based on each *S*'s cardiac rate prior to that response. Such a correction is designed to take account of the fact that cardiac acceleration to a constant stimulus is negatively correlated with the prestimulus rate. The correction was derived by correlating prestimulus level and response to shock for 20 *Ss* over 10 trials at 90-sec. intervals. These measurements were made in another experimental situation, and with different *Ss* from those used in the present study. The product-moment correlation between these 200 pairs of values was $-.67$ and the regression of response on prestimulus level had a best-fit line of $Y' = -.12X + 29.9$.

This line was broken into segments to allow an additive correction. For each 5 BPM the prestimulus level was above 60 BPM, 0.6 BPM was added to the measured response. This provided a very close approximation to the regression line. Both pretest responses and conditioning trial responses were corrected on the basis of this scale. The correction was applied only to the mean measurements for the various time periods, and is not applicable to the second-by-second analysis.

Because there were no progressive changes in the mean prestimulus rate for the total group as a function of trials, the correction did not materially alter the general shape of the conditioning curve plotted from uncorrected data. The correction however, did serve to reduce variability both among *Ss* and across trials.

TABLE 1

DIFFERENCES IN RESPIRATION AND CARDIAC RESPONSE BETWEEN MEANS FOR THE LAST THREE PRETESTS AND THE LAST THREE CONDITIONING TRIALS

Measure	Respiration ($N = 41$)*		Cardiac Response ($N = 44$)
	Rate (Cycles/ CS-UCS Interval)	Depth (Millimeters Deflection)	Beats
Mean diff.	+.006	+.20	+3.97
$\sigma_{\text{Mean diff.}}$	.04	.13	.57
CR	.15	1.54	6.96
P	>.50	>.10	<.01

* For three *Ss*, respiration records were illegible on one or more trials.

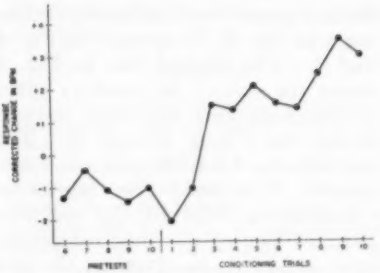


FIG. 3. Acquisition curve for a conditioned cardiac response employing buzzer, shock, and respiratory control. Each point is the mean for 44 *Ss*.

Results

The respiration rate was analyzed in terms of cycles per CS-UCS interval and the depth in terms of millimeters deflection from peak to trough on the original recordings. There were no significant changes in respiration from the last three pretests to the last three conditioning trials as can be seen in the first two columns of Table 1. Neither mean difference is significantly different from zero.

The net results for the heart rate are shown in Fig. 3 and in the third column of Table 1. The curve indicates that the effect of the buzzer and the controlled respiration was a net decrease in the heart rate of 1.1 BPM as an average for the last three pretests. With the institution of the shock at the termination of the buzzer, the response changed from a deceleration to an increasingly marked acceleration reaching a peak of 2.9 BPM as a mean for the last three conditioning trials. This change is significant well beyond the .01 level (cf. Table 1). The form of the acquisition curve approximates a negatively accelerated function of trials.

The second-by-second analysis of

the last pretest and the last conditioning trial for all *Ss* appears in Fig. 4 and 5. The dashed line in Fig. 4 shows the smooth deceleratory effect of the controlled breathing pattern during the CS on Pretest 10. The rate falls by 4.7 BPM over the 7-sec. period. The solid line represents Conditioning Trial 10 and indicates that at CS onset, the smooth deceleration observed in Pretest 10 was replaced by a steady acceleration of 8 BPM reaching this peak 4 sec. after CS onset and followed by a decline of 10 BPM for 3 sec. until the onset of the UCS. At that point the heart accelerated rapidly by 27.1 BPM, reaching this peak 3 sec. after UCS onset. It then declined slightly, but 7 sec. later was still more than 20 BPM above the preshock level.

In Fig. 5, the solid line shows the cardiac activity during Conditioning Trial 10 plotted on a baseline of the second-by-second activity observed in Pretest 10. Each point is the difference between the rates at each second of these two comparison trials. The response attributable to conditioning may be described as an acceleration, reaching its peak 4 sec. after CS onset, followed by a decline until UCS onset. At no point on this corrected curve does the rate during the CR fall below the pre-CS rate.

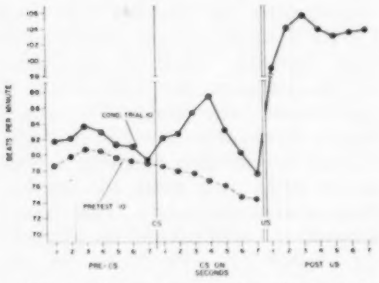


FIG. 4. Second-by-second cardiac activity on the last pretest and the last conditioning trial for all *Ss* with respiration held constant.

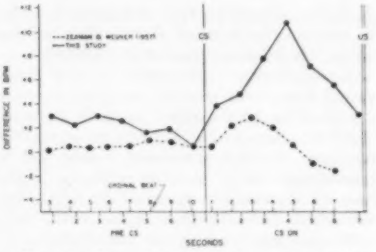


FIG. 5. Second-by-second cardiac activity on the last conditioning trial, with respiration held constant, corrected for cardiac activity on the last pretest. Each point is the difference between the rate on the last pretest and the rate on the last conditioning trial. Superimposed is a comparable curve with respiration uncontrolled, adapted from Zeaman and Wegner (1957, p. 131).

The dashed line is a comparable curve of response presented by Zeaman & Wegner (1957, p. 131) as a beat-by-beat analysis of the records of 63 *Ss* studied in several different conditioning situations with a constant 6-sec. CS-UCS interval. Their curve was raised along the ordinate so that the two curves would be matched at CS onset. The differences between these two curves will be considered in the discussion.

Discussion

The results of these experiments can be addressed to three questions. First, is cardiac conditioning in human *Ss* simply an artifact of respiratory activity? The general issue of artifact in conditioning studies has been raised by Smith (1954) and has been handled by Black (1956) through the use of paralytic drugs on dogs. The present study demonstrates that in humans, cardiac conditioning is indeed more than simply an accompaniment of respiratory changes. In the absence of respiratory changes, progressive and predictable changes in cardiac activity were observed as a function of trials.

The second issue is the extent to which previously reported conditioned cardiac responses are seriously influenced by

respiratory activity. The present study has shown respiratory activity to be an important and controllable variable. In Fig. 5 the Zeaman and Wegner (1957) curve differs from the curve produced in the present study in respects which one would expect if a direct cardiac conditioning effect were superimposed on a fluctuating respiration-produced baseline similar to that shown in Fig. 2 rather than on a smooth one similar to the dashed curve in Fig. 4 produced by controlled respiration. The Zeaman and Wegner (1957) curve has a shorter time to maximum and a more marked deceleratory phase. The different amplitudes of the two curves in Fig. 5 may be accounted for by the different intensities of noxious stimulus. Zeaman and Wegner (1957) have suggested that previously reported forms of response may have been multiply determined by direct cardiac conditioning and respiratory activity. The present study strongly suggests that this is the case.

The third issue concerns the mechanisms by which a conditioned cardiac response is acquired. The acquisition curve in the present study approximates an increasing monotonic function of trials. This would be predicted by both drive reduction and stimulus-substitution views of conditioning. It is the precise beat-by-beat form of a single, well-established response which would be expected to tell whether cardiac activity during the CS is more similar to that at UCS onset (stimulus substitution) or at UCS termination (drive-reduction). Zeaman & Wegner (1954, 1957) and Wegner and Zeaman (1958) have explored this question, but in none of their experiments was respiration controlled. Similarly, the present study casts no light on this issue inasmuch as the controlled respiration during the CS inevitably broke down upon UCS presentation, and it is impossible to tell what is the "real" form of the UCR. The control of both respiratory activity and gross bodily movement during all phases of such stress situations poses formidable technical problems with human *Ss*, but it appears that these must be solved prior to further illumination of this issue.

Summary

Previous studies of cardiac conditioning have yielded conflicting results. Two studies were performed to clarify some of the discrepancies. Experiment I investigated the effects of various patterns of respiratory activity on cardiac activity. Reliable cardiac patterns were found to be associated with particular respiratory patterns. Thus, the cardiac accompaniments of conditioned respiratory activities may be mistaken for direct cardiac conditioning. Experiment II utilized respiratory control in a study of cardiac conditioning, pairing a buzzer and an electric shock. A conditioned cardiac response, independent of respiration, was observed. The form of the conditioned cardiac response and the acquisition curve of the response were plotted and compared to earlier results. The implications of these results for further studies of cardiac conditioning and for theoretical interpretations of cardiac conditioning were discussed.

References

Black, A. Extinction of avoidance responses under curare. Unpublished doctoral dissertation, Harvard University, 1956.

Huttenlocher, J., & Westcott, M. Some empirical relationships between respiration and heart rate. *Amer. Psychologist*, 1957, **12**, 414. (Abstract)

Notterman, J., Schoenfeld, W., & Bersh, P. Conditioned heart rate response in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45**, 1–8.

Smith, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, **61**, 217–225.

Wegner, N., & Zeaman, D. Strength of cardiac conditioned responses with varying unconditioned stimulus durations. *Psychol. Rev.*, 1958, **65**, 238–241.

Zeaman, D., Deane, G., & Wegner, N. Amplitude and latency characteristics of the conditioned heart response. *J. Psychol.*, 1954, **38**, 235–250.

Zeaman, D., & Wegner, N. The role of drive reduction in the classical conditioning of an autonomically mediated response. *J. exp. Psychol.*, 1954, **48**, 349–354.

Zeaman, D., & Wegner, N. A further test of the role of drive reduction in human cardiac conditioning. *J. Psychol.*, 1957, **43**, 125–133.

(Received March 23, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 360–364

RESPONSE LATENCIES PRODUCED BY MASSED AND SPACED LEARNING OF A PAIRED-ASSOCIATES LIST

JOHN BROWN AND M. HUDA

Birkbeck College, University of London

In studies of the learning of a serial list or a list of paired associates, it is customary to control the rate of presentation and thus to limit the response period, i.e., the time available to *S* for each anticipation or response. The response period is usually about 2 sec. This is a comparatively short period for a response which involves recall. For example, one investigator found that a number of correct responses were made in the second half of a 4-sec. response period (Osgood, 1948). Consequently the results obtained when the response period is limited may often reflect changes in response latencies rather than changes in the number of responses available, i.e., the number of responses which would be available if the response period were longer. Response latencies could change in several ways, without any change in the number of responses available. Thus the distribution of latencies might change in either mean, variance or shape and each of these changes could lead to a greater incidence of high latency responses. The possibility arises that the observed effects of massed practice on serial and paired-associate learning are due primarily to a tendency for response latencies to be higher under massed practice than under spaced. Thus, suppose the criterion of learning is a trial with all responses correct. If some of the latencies under massed practice exceed the response period on the trial at which the criterion is reached under spaced practice, then additional trials would be needed under massed practice. Similarly, reminiscence may reflect not so much an increase in the number of responses available as a decrease in response latencies during retention. Decrease in latencies will tend to produce reminiscence if some responses are available on the last learning trial but could not be made quickly enough. (This is a slight oversimplification since under the control condition for reminiscence a test trial immediately follows the last learning trial. What counts therefore is the latencies on this test trial.) Naturally changes in response latencies will not affect the number of correct responses if all latencies are less than the response period. The finding that decreasing the rate of presentation decreases the effect of massing (Hovland, 1938b) and also decreases the amount of reminiscence obtained (Hovland, 1938a) is therefore suggestive evidence for the importance of latencies.

The main object of the present experiment was to compare the response latencies produced by massed and spaced practice. A second object was to see whether a retention interval of 15 min. differentially affects the latencies associated with massed and spaced practice. Ward (1937) obtained some evidence that following massed practice, response latencies may be lower after 2 min. or after 5 min. than after 6 sec. (the control) but Melton and Stone (1942) found a trend in the opposite direction for these same time intervals. In both

these investigations serial learning was studied. With this type of learning, the order of the responses is fixed so that *S* can prepare himself in advance for the next response. This means that the observed response latency is an equivocal measure of the time taken to recall the response. Accordingly paired-associate learning was chosen for the present study, since the order of the pairs and hence of the responses can be varied.

Method

Stimulus materials.—The learning list consisted of 10 pairs of adjectives. The stimulus words were 10 words of medium similarity used by Underwood (1953). There was low similarity among the response words which were paired so as to eliminate any obvious relations of sound or meaning between the members of each pair.

Apparatus.—The list was typed in capitals around the periphery of a white cardboard disc in three different orders. Each word or pair of words lay on a different radius. The disc was clamped on to an electromagnetic device controlled by a timer which turned the disc one step every 2 sec. In front of the disc was a screen with a viewing aperture. The stimulus word alone appeared in the aperture for 2 sec. and was followed by the stimulus word and the response word for 2 sec. The *S*'s responses were recorded via a microphone on to a magnetic tape. The microphone was placed so that it picked up a click made by the electromagnetic device each time it turned the disc forward.

Latency measurement.—Latencies were measured by two methods. The first method involved direct measurement of the distance on the tape between the click which indicated the onset of the stimulus and the start of *S*'s response. Their locations were found by moving the tape to and fro past the playback head as in tape editing. The second method consisted in playing back the tape into a pen recorder and measuring the latencies from the visual record. Latencies measured by the two methods never differed by more than .1 sec.; any difference was always in the direction of slightly shorter latency with the first method, presumably because of inertia in the pen recorder. The first measure was employed with all 20 *S*s and the second method with 10 *S*s in each group.

Procedure.—The intertrial interval was 4 sec. under the massed condition and 40 sec. under the distributed condition. Under the distributed practice the rest interval was filled with a cancellation task. Learning was carried to a criterion of three successive error-free trials. On completion of learning *S*s were engaged in casual conversation and care was taken to ensure that *S* carried the main burden of the conversation. A single recall trial under exactly the same conditions as learning was given 15 min. after the completion of learning. At the outset of the experiment the nature of the learning task was explained to each *S* and a check was made to confirm that the instructions were understood.

Subjects.—Forty university students of both sexes, 20 under each condition, were successively assigned to one condition or the other in random order. They were all naive to this kind of verbal learning.

Results

The mean response latency on the three criterion trials was 1.562 sec. for the massed group and 1.374 sec. for the spaced group. The difference is highly significant ($t = 3.39$, $df = 38$, $P < .002$). The mean number of trials to reach the criterion was 24.5 for the massed group and 19.5 for the spaced group, including the first of the criterion trials. The difference is not quite significant at the 5% level on a t test ($t = 1.93$, $df = 38$). This test is not really appropriate, however, since several *S*s required many more trials than the average. When trials-to-learn in the two groups were compared by the Mann-Whitney non-parametric test, the difference became just significant at the 5% level.

Table 1 shows the latencies, for responses correctly made, at various stages of learning and on the retention trials for both the massed and spaced groups. It also shows the mean number of trials needed to reach each stage of learning by the two groups. Stages I and II represent the trials on which *S*s first attained at least 5/10 correct and at least 8/10 correct, respectively. Stage III is quoted

TABLE 1

MEAN RESPONSE LATENCIES AND TRIALS AT DIFFERENT STAGES OF LEARNING AND RECALL

Stage of Learning or Recall	Massed Group		Spaced Group	
	Mean Latency (Sec.)	Mean Trials	Mean Latency (Sec.)	Mean Trials
Learning				
I. ≧5/10 correct	1.600	7.9	1.560	7.2
II. ≧8/10 correct	1.615	15.9	1.485	14.0
III. 10/10 correct	1.565	21.7	1.405	16.5
IV. 1st criterion trial	1.579	24.5	1.390	19.5
V. 3rd criterion trial	1.557		1.364	
Recall trial	1.510		1.375	

in order to facilitate comparison with experiments in which the criterion of learning is a single error-free trial. Potentially, Stage II coincides with the first of the 3 criterion trials but 16 *Ss* in the massed group and 15 *Ss* in the spaced group made further errors after this stage.

Examination of Table 1 shows that the difference between the massed and spaced groups emerges as learning proceeds, as we would expect. Thus at Stage I, the difference in latencies is only .040 sec., and the difference in trials is less than a single trial: neither difference is statistically significant. The difference in latencies first becomes significant at Stage II ($(t = 2.19, df = 38, P < .05$) and in trials at Stage III ($P < .05$ on a Mann-Whitney test). The latency fell by .170 sec. between Stages I and IV in the spaced group but by only .021 sec. in the massed group. The decrease for the spaced group is significant ($t = 3.49, df = 38, P < .002$) but not the decrease for the massed group ($t < 1$). When the two decreases are compared, the difference (.149 sec.) is significant ($t = 2.19, df = 38, P < .05$).

Turning to data concerning retention, the mean latency on the recall trial was 1.510 sec. for the massed group and 1.375 sec. for the spaced group and the difference is significant ($t = 2.10, df = 38, P < .05$). This difference is a little smaller than the corresponding difference on the criterion trials but it is not significantly smaller as shown by comparing the scores for the difference between the mean latency on the criterion trials and the mean latency on the recall trial for the *Ss* of the two groups ($t = 1.78, df = 38, .05 < P < .1$). The mean latency on the recall trial was almost identical with the mean latency on the third criterion trial for the spaced group but was .047 sec. lower for the massed group. Although this is only a small decrease, it is just significant ($t = 2.13, df = 19, P < .05$). There was a nonsignificant trend in favor of better retention with the spaced group: only 3 *Ss* of this group made errors on the recall trial as compared with 9 *Ss* of the massed group ($\chi^2 = 2.97$). Expressed as percentages of pairs correct, recall was 93.5% in the massed group and 98% in the spaced group.

DISCUSSION

The massing of practice trials may have more than one effect on learning. The results show clearly that massed practice produces higher response latencies than spaced practice. In a very different field, Ratner and Miller (1959) have recently reported that when conditioning trials are massed, response latencies in earthworms are higher than when the trials are spaced. The finding that massed practice results in higher latencies helps to explain a number of results reported in the literature. Any factor which increases response latencies will increase the chance that some of the latencies under massed practice will exceed the response period on the trial at which the criterion is reached under spaced practice. In making this statement, a criterion in terms of the proportion of correct responses is assumed.

Factors likely to increase response latency include high stimulus similarity, low meaningfulness of responses and fatigue induced by performance of a difficult mental task immediately before learning. Each of these factors has been found to increase the effect of massed practice on trials-to-learn: high intralist similarity and low meaningfulness by, for example, Underwood and Richardson (1958) and Underwood and Schulz (1959) and the effect of fatigue by Hovland and Kurtz (1951). Underwood and Schulz (1959) have suggested that massed practice increases trials-to-learn only when some degree of response competition is present. Whether or not this view is correct, the degree of response competition is likely to be important for two reasons. First, it is likely to increase latencies and, as pointed out earlier, higher latencies under massing will not increase trials-to-learn if all latencies are within the response period. Second, response competition may increase latencies under massing more than it increases latencies under spacing. This would further raise the probability of massed practice increasing trials-to-learn. Massed practice has been found to increase trials-to-learn less often with paired associates learning than with serial learning. This might be due to a greater incidence of high response latencies during the learning of a serial list, since the formation of remote associations is likely to cause response competition and high latencies in the middle of the list.

It has often been suggested that massed practice tends to increase trials-to-reach-criterion because of the accumulation of some form of inhibition which depresses performance and which is able to dissipate in the rest intervals of spaced practice. The finding that massed practice results in longer latencies even when the criterion is reached might be regarded as further evidence for this view since the presence of inhibition would presumably increase latency. But if it were correct a 15-min. retention interval should allow the inhibition causing the long latencies after massed practice to dissipate, thus making the latencies for the massed and spaced group equal. In fact the difference between the latencies for the two groups fell from .185 sec. to .135 sec. only, a change which was not statistically significant (the 5% confidence limits for the change are +.004 and +.266 sec.). The fall in latencies between the last criterion trial and the recall trial for the massed group although statistically significant was small (.047 sec.). These findings suggest that accumulation of an inhibition affecting performance is not a major factor in raising latencies under massed practice. It might be argued that the 15-min. retention interval was inadequate to allow the inhibition to dissipate. But unless the inhibition dissipates fairly rapidly, dissipation of inhibition during the 40-sec. intertrial interval under spaced practice will not be sufficient to account for the lower latencies produced by this type of practice. An alternative interpretation of the higher latencies associated with massing is that learning is itself different under massing. Learning usually involves an active search for meaningful relations within the material and also for helpful associations. This active side of the learning process may be impaired by a fatigue factor under massing. Such impairment would not only slow learning but might also alter its character. Higher latencies under massing would, according to this view, remain higher (unless modified by further learning) and would not diminish with the dissipation of fatigue. Underwood reached similar conclusions after he had failed to find a reminiscence effect when a rest interval was interpolated during massed practice: "it does not now seem justifiable to postulate a rapidly dissipating inhibition which depresses performance and does not influence associative strength" (Underwood, 1957, p. 143).

The results showed that, in the spaced group, the mean response latency fell significantly between Stage I of learning (at least 5/10 correct) and the first criterion trial but, in the massed group, it remained almost stationary. These

findings should be interpreted with caution. In general we might expect practice to reduce latencies. However, when a 5/10 stage is compared with a 10/10 stage, at least two possible factors which may influence the observed change in latency should be taken into account. First, at the 5/10 stage *S* is presumably making only those responses which are relatively easy for him and which may therefore have a naturally lower latency than the more difficult responses he has yet to acquire. Second, at the 5/10 stage the number of alternative responses between which the *S* chooses when a stimulus appears may be lower than at a 10/10 stage, since at the 5/10 stage he may ignore some of the pairs altogether. This would also tend to reduce latencies at the 5/10 stage.

With paired-associate learning, massed practice does not often produce a statistically significant increase in trials-to-learn, although the trend is nearly always in this direction (Underwood, 1953). In the present experiment, the increase was significant but only at the 5% level. However, the actual size of the increase was quite large (28%). Thus the reason why the increase is often not significant is probably that the variability in trials-to-learn from *S* to *S* is high rather than that the increase itself is negligible. It is plausible to attribute the observed increase in trials-to-learn for the massed group to increased response latencies, in the manner explained previously. This hypothesis could be further explored by introducing a trial during learning with a drastically reduced rate of presentation. If the hypothesis is correct, as many correct responses should be made by massed practice *S*s, although the response latencies for the former would be higher.

Summary

Two groups of 20 *S*s learned a list of paired adjectives to a criterion of three successive error-free trials. The rate of presentation was 4 sec. per pair. For one group the intertrial interval was 4 sec. and for the other group it was 40 sec. A recall trial was given 15 min. after learning. The following results were obtained: (*a*) the spaced learning group required significantly fewer trials ($P < .05$) to reach the criterion; (*b*) the response latencies were significantly ($P < .002$) higher on the criterion trials for the massed group; (*c*) this applied also on the recall trial ($P < .05$); (*d*) nevertheless response latencies fell significantly ($P < .05$) between the third criterion trial and the recall trial only in the case of the massed group; (*e*) there was a nonsignificant trend in favor of better retention by the spaced group.

References

Hovland, C. I. Experimental studies in rote-learning theory: II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22,** 338–353. (a)

Hovland, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23,** 172–190. (b)

Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory: IX. Influence of work decrement factor on verbal learning. *J. exp. Psychol.*, 1951, **42,** 265–272.

Melton, A. W., & Stone, G. R. The retention of serial lists of adjectives over short time intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30,** 295–310.

Osgood, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38,** 132–154.

Ratner, S. C., & Miller, K. R. Effects of spacing of training and ganglia removal on conditioning in earthworms. *J. comp. physiol. Psychol.*, 1959, **52,** 667.

Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45,** 143–149.

Underwood, B. J. Studies of distributed practice: XVI. Some evidence of the nature of the inhibition involved in massed learning of verbal materials. *J. exp. Psychol.*, 1957, **54,** 139–143.

Underwood, B. J., & Richardson, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56,** 213–219.

Underwood, B. J., & Schulz, R. W. Studies of distributed practice: XIX. The influence of intralist similarity with lists of low meaningfulness. *J. exp. Psychol.*, 1959, **58,** 106–110.

Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**(4, Whole No. 220).

(Received March 11, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 365-367

ANALYSIS OF DOUBLE ALTERNATION IN TERMS OF PATTERNS OF STIMULI AND RESPONSES

JAMES W. SCHOONARD AND FRANK RESTLE

Michigan State University

This paper reports the results of an experiment in which college students were taught double alternation by two methods.

If *S* learns a single response to a fixed stimulus situation, R R R R R, it is possible that a simple stimulus-response association has been formed. Alternating two responses in the same situation, R L R L R L, may be attributed to some type of response inhibition or stimulus satiation, or to discrimination learning based on simple stimulus traces from the previous response. In double alternation, R R L L R R L L, no such simple mechanisms will explain the behavior, for responses must be based on the outcomes of at least two previous trials.

The nature of the process which mediates double alternation is not well understood. One could argue (*a*) that *S* perceives and remembers a sequence of previous reinforcing events, basing responses in a discriminative fashion on such patterns, (*b*) that past responses are remembered and used as discriminative cues, or (*c*) that individual choices unite to form a larger response chain of "double alternating."

In an attempt to separate these possibilities experimentally, the same double alternation problem was taught in two ways. One training method favored the use of stimulus patterns and the other favored the use of response patterns. The physical environment was exactly the same at each response as in the "temporal maze" for animals (Hunter & Nagge, 1931) so that behavior was not controlled by any extraneous stimuli. The *S* was informed of the correct response no matter what his choice, giving full knowledge of results. Since the procedure used was similar in some respects to "verbal conditioning" (Estes & Straughan, 1954), the results can be related to current mathematical models of such learning.

METHOD

Subjects.—The *S*s were 40 students drawn from an introductory psychology course at Michigan State University, who volunteered but received some academic credit for experimental work. They were assigned alternately to experimental groups in the order of their appearance, except for an attempt to equalize the sex ratio of the groups. The data from 12 other *S*s were discarded owing to (*a*) misunderstood instructions, (*b*) an error in procedure, or (*c*) the fact that *S* made no errors and hence (for reasons explained below) could not be assigned to either group.

Procedure.—The *S* and *E* were separated by a partition. On *S*'s side was a device which produced a click every 6 sec. He was told to respond to each click by saying "check" or "plus" immediately after the click. The *E* then indicated the correct response by saying either "check" or "plus." Responses by *S* (represented here by P and C) were to be predictions of what *E* would say next (here represented by p and c).

For Group S (the stimulus-pattern group) the events (*E*'s replies) followed a fixed double alternation pattern, p p c c p p etc. Thus if *S* tried to form a pattern of past events his task should be easy. Since the events were not contingent on *S*'s responses, he might be reinforced for response patterns other than double alternation. Furthermore, he might be nonreinforced even if he produced a pattern such as P P C C P, provided it was out of phase with the sequence of events.

For Group R (the response-pattern group)

the events were made contingent on *S*'s responses, so that any double alternating pattern of responses would be correct and any deviation wrong. For example, if *S* said P on Trial 5 and P on Trial 6, *E* would say c on Trial 7. This procedure reinforces the sequence P P C if it occurs, for that is part of a double alternation sequence, and nonreinforces a deviation such as P P P if it occurs. This system does not determine the reinforcement for the first two responses, which were always called correct. With this procedure, *S* was consistently reinforced whenever he began a double alternating pattern of responses, and consistently nonreinforced any time he produced a conflicting pattern of responses. An *S* who never erred could belong to either group, and his data must be discarded from the comparative analysis.

Six consecutive double-alternations (12 correct responses) were taken as the learning criterion. If *S* did not attain this criterion in 100 trials, training was terminated. An *S*'s performance was summarized by the trial on which he last made an incorrect response.

Results

The stimulus-pattern problem was clearly easier than the response-pattern problem. All *S*s in Group S solved their problem, with a median trial-of-last-error of 10. Only 9 of the 20 *S*s in Group R solved their problem. The median trial-of-last-error of Group R was 95 (on the basis of 100 trials maximum). A simple median test indicates that the difference is significant beyond the .001 level. The data are summarized in Table 1.

TABLE 1

Trial of Last Error for Groups S and R

Group	Trial of Last Error			
	Mean	Median	*SD*	Range
S	15.6	10.0	12.3	3–49
R	67.5	95.0	[a]	7–100

[a] The restriction in range of scores caused by terminating the training at 100 trials makes this value meaningless.

The irregular responses of *S*s in Group R before solution sometimes gave rise to sequences of events which double alternated temporarily. Thus the *S*s of Group R had pieces of the stimulus-pattern problem from time to time. The stimulus-pattern problem is learned so rapidly that one may conjecture that *S*s solved the response-pattern problem only by solving the stimulus-pattern problem when it occured accidentally. If this is so we should find that *S*s who solved the response-pattern problem had a relatively high frequency of accidental double-alternating runs of events before solution, and the nonsolvers should have had a relatively low frequency of such runs.

To test this possibility, a run of events was defined as five consecutive events which double alternate, such as p p c c p or c p p c c, etc. This run of five seems just long enough to characterize double alternation. A run of seven events like p p c c p p c was counted as three (overlapping) runs. If, contrary to our conjecture, learning is independent of the number of runs, then the number of runs an *S* experienced before solution should be about proportional to the number of trials before solution. However, the first four trials of the problem cannot mark the end of a run, so we deal with the number of trials minus four.

Let the trial of last error be n and the number of runs seen before solution be r. Then $r/(n-4)$ is an index of the proportion of runs seen by *S*. The mean of this index for the 9 solvers in Group R was .208 ($SD = .116$). For nonsolvers the mean was .149 ($SD = .058$). The difference is favorable to our conjecture but it might reasonably be attributed to sampling variation ($t = 1.47$).

Discussion

Evidently, the sequence of events is a salient and effective guide to double alternating, since Group S learned very quickly. It is equally clear that double alternating chains of responses are not easily formed even with perfectly consistent reinforcement. The fact that some *S*s in Group R solved the problem within 100 trials indicates either that they were finally "shaped" by reinforcement or that they remembered enough of their past responses and reinforcing events to constitute usable discriminative cues. It is even possible, though not proved, that the solvers in Group R succeeded by noticing runs of double alternation in events, responding accordingly, and thereby causing the pattern of events to remain orderly.

The quick learning of Group S shows that college students can easily take advantage of patterns of events in this sort of verbal conditioning experiment. Thus our very simple procedure supports the findings of Hake & Hyman (1953), Goodnow (1955), Engler (1958), Nicks (1959), and Anderson (1960), all of whom have studied patterning in verbal conditioning situations. With a determinate sequence such patterns are formed quickly and fixated solidly, in contrast to the slower learning and more variable response to partly random sequences.

Our tentative conclusion is that double alternation problems require *S* to remember previous events at least far enough back to discern the pattern. The animal studies (Gellermann, 1931a, 1931b, 1931c; Hunter & Nagge, 1931) can be interpreted to mean that rats have great difficulty with this sort of immediate memory task as they do in delayed-response experiments, and that higher animals are typically more skillful at remembering past events.

Summary

Two groups of 20 college students each were taught double alternation in a verbal conditioning situation, one group being given a fixed sequence of events and the other group being selectively reinforced for double alternating response-chains. The fixed sequence of events was solved much more rapidly. The results suggest that college students can easily remember a sequence of several events and base their responses on patterns observed.

References

Anderson, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59,** 73–93.

Engler, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55,** 303–317.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47,** 225–234.

Gellermann, L. W. The double alternation problem: I. The behavior of monkeys in a double alternation temporal maze. *J. genet. Psychol.*, 1931, **39,** 50–72. (a)

Gellermann, L. W. The double alternation problem: II. The behavior of children and human adults in a double alternation temporal maze. *J. genet. Psychol.*, 1931, **39,** 197–226. (b)

Gellermann, L. W. The double alternation problem: III. The behavior of monkeys in a double alternation box-apparatus. *J. genet. Psychol.*, 1931, **39,** 359–392. (c)

Goodnow, J. J. Response-sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68,** 624–630.

Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45,** 64–74.

Hunter, W. S. The sensory control of the maze habit in the white rat. *J. genet. Psychol.*, 1929, **36,** 505–537.

Hunter, W. S., & Nagge, J. W. The white rat and the double alternation temporal maze. *J. genet. Psychol.*, 1931, **39,** 303–319.

Nicks, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57,** 105–114.

(Received March 17, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 368-375

JUDGMENTS OF SURFACE ILLUMINATION AND LIGHTNESS[1]

JACOB BECK

University of Pennsylvania

The hypothesis that judgments of lightness and illumination are coupled psychologically was originally formulated by Helmholtz (1925) and has been frequently restated since (Koffka, 1935; Woodworth & Schlosberg, 1954). Though of considerable heuristic value, the hypothesized relationships between these two response variables fails to be predictive in many instances (Henneman, 1935; MacLeod, 1940, 1947; Thouless, 1932). Beck (1959) proposed an alternative analysis based upon independently specifying the correlates for the judged lightness and illumination of a surface. He reported on the judgments of lightness and illumination of a single surface consisting of a clearly perceptible pattern composed of two different reflectances. Judgments of illumination were found to be strongly influenced by the higher of the two light intensities reflected from the surface while judgments of lightness were a function of the relative intensities.[2] This study suggested the possibility of specifying those conditions where changes in either the objective illumination or the reflectances of the surfaces in the visual field would change either judgments of lightness, illumination, or both based upon the interrelations of the stimulus correlates involved. The purposes of the present experiments were, first, to discover for different kinds of surface textures, with and without a background, the principal factors governing *O*s' judgments of surface illuminance and, second, to examine the relationship between judgments of surface, illumination and lightness for these conditions.

The experiments used binocular matching to investigate the judgments of illumination and lightness produced by the four surface textures shown in Fig. 1. The *O*s' illumination matches are compared with one or more of the following measures according to the experimental conditions: the illuminance of the comparison field which equates (*a*) the average brightness, (*b*) the points of maximum brightness, (*c*) the background brightnesses and (*d*) the illuminance of the standard and comparison surfaces.

METHOD

Apparatus

The apparatus was similar to that used in Exp. IV of Beck (1959). The standard and comparison surfaces were placed in two large adjacent light proof compartments from whose front side two rectangular windows had been cut. A partition with a chin rest attached to one end was placed perpendicular to the front side of the compartments and between the two viewing windows. The *O* when placing his chin in the chin rest could see the standard and comparison surfaces looking with one eye through each window.

Illumination on the standard and com-

[1] This research was supported by Grant B-1876 from the United States Public Health Service. The author is indebted to Howard Eisman, Doreen Rothman, Lester Phares, and Francis Smith for the collection of much of the data.

[2] The correlation between *O*s' illumination matches and the higher light intensity in the four experiments was .84. Changing the illuminance had little affect on *O*s' judgments of lightness indicating it to be a function of the relative intensities reflected from the surfaces.

parison surfaces was provided by two light boxes, each consisting of 100 6-w. bulbs, placed directly in front of the surfaces below the line of sight. The *O* controlled the illumination on the comparison surface by means of a telephone switch which operated a motor causing a beveled plate to move against a series of microswitches controlling the individual bulbs of the light box. The light box was wired so that the illuminance could be varied in 40 steps from complete darkness to a maximum. The sounds of the opening and closing of the microswitches were muffled by padding and masked through the use of a metronome. The *E*'s control apparatus was placed on a table located at *O*'s left. A specially prepared scale calibrated with respect to a pointer attached to the beveled plate enabled *E* to directly record *O*s' settings of the switch. The *E* was able to independently change the setting of the switch and control the light boxes illuminating the standard and comparison surfaces.

Voltage regulators were used to keep the line voltage constant. All measurements of illumination were taken with a Spectra brightness spot meter.

Procedure

The experiments required *O*s to adjust the illuminance on a comparison surface until its illumination was judged as the same as that of a standard surface while simultaneously viewing each with monocular vision and a motionless head in a completely dark room. The procedure and instructions were similar to those used by Beck (1959). Each *O* was seated in the darkroom before the chin rest and read the instructions which included: (*a*) the distinction between the lightness of a surface and its illumination; (*b*) operating detail concerning the telephone switch and the method of adjustment; (*c*) that his task was to adjust the illumination incident on the comparison surface until it appeared equal to the illumination of the standard; (*d*) that he should look at the surfaces naturally and make his judgments solely on their visual appearance; and (*e*) that there was no time limit for the task. Two levels of illumination were used in each of the experiments, *O* making 10 separate matches at each intensity level. The *E* alternately set the illumination on the comparison surface to a point obviously too low or too high. Each *O* was dark adapted for 3 min. before the experiment, and approximately 2 min. elapsed between each of the 10 matches made. The *O*s were given two practice judgments.

At the conclusion of the illumination matches, the illumination on the standard surface remained at the value used in the experiment and the illumination on the comparison surface was set at the median *O*'s matches. The *O* was then asked to match the lightness of the comparison and standard surfaces with a chart of 12 Hering grays which varied in reflectance from 8% to 78%. The chart was illuminated by a 25-w. bulb which was turned on after *O* viewed the stimulus surface. Two different charts were prepared, each consisting of the same 12 gray samples arranged in a different order. A different chart was used with each of the two illumination levels of the standard.

The brightnesses of the target and background surfaces in the standard and comparison fields in the different experiments were measured with the spectra brightness spot meter. For the comparison field, the intensities in foot-Lamberts corresponding to each of the 40 switch positions were determined. To provide a common base for comparisons, all illumination measurements are presented in terms of the average brightness of the target surfaces. Table 1 shows the average brightness of the standard surface for the two illumination levels used, and that of the comparison surface corresponding to the medians of *O*s' illumination matches. Similarly, the other illumination measures, such as the settings for the comparison field which equate the illuminance of the target surfaces or the brightnesses of their backgrounds, are expressed in terms of the average brightness of the comparison surface. Regular checks were made for bulb aging in the light boxes illuminating the standard and comparison surfaces.

Observers

The *O*s were students at the University of Pennsylvania. All had normal vision with correction and were naive regarding the experiments. Fifteen *O*s were used in each experiment, except for Exp. VII where only 12 were used.

Conditions

The two experimental variables are surface texture and the presence of a background. Optically, the variables manipulated are the brightness gradients present within the surface and the presence of a surround of a higher brightness than that of the surface. Except for Exp. III, where the standard and comparison surfaces were both of a uniform brightness, the experiments were all conducted with surfaces possessing a microstructure.

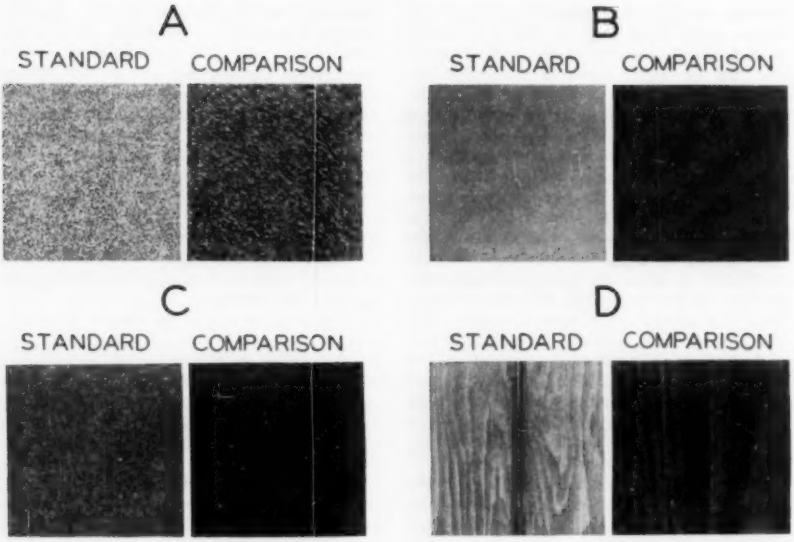


FIG. 1. Photographs of the standard and comparison surfaces employed in the different experiments.

They thus differ from the "speckled" and "striped" surfaces previously investigated (Beck, 1959) which consisted of a clearly perceptible pattern of two different uniform brightnesses.

Experiment I.—The standard and comparison surfaces used in Exp. I are shown as A in Fig. 1. They consisted of black, dark gray and gray Zipatone patterns placed over one another on a white backing. The surfaces were placed 107 in. from *O*, and set so that he could see approximately $6\frac{1}{4} \times 6\frac{1}{4}$ in. areas of the standard and comparison surfaces looking with one eye through each viewing window. The *O* was able to see only the target surfaces. A small fixation point was placed in the center of each of the areas.

The average reflectance of the standard surface was 50% and of the comparison surface 28%. The points of maximum brightness coincided with the clear portions of the Zipatone sheets and were the same for the standard and comparison surfaces. Two levels of illumination were used; the average brightness of the standard surface was set at 2.7 and 6.8 ft.-L. To equate the points of maximum brightness of the standard and comparison surfaces, the average brightness of the comparison surface had to be set at 1.54 and 3.88 ft.-L. These values were photometrically checked by measuring the brightnesses of the white sheets employed as backing for the Zipatone patterns used for the standard and comparison surfaces. Since the maximum reflectances of the two surfaces were the same, a match equating the points of maximum brightness would also equate the illuminance of the standard and comparison surfaces.

Experiment II.—The standard and comparison surfaces were the same as those of Exp. I. They were now placed 90 in. from *O* and set so that above and to the right of each surface a white background situated 21 in. behind the surfaces was visible. The areas of the standard and comparison surfaces visible to *O* were approximately 11 in. × 5 in. The *O*s were able to see only the target surfaces and their backgrounds.

Two levels of illumination were used: the average brightness of the standard surface being set at 2.4 ft.-L. and 9.3 ft.-L. The brightness of the background behind the standard surface for the two illumination levels was 2 ft.-L. and 7.8 ft.-L. To equate the brightnesses of the backgrounds behind the standard and comparison surface, the average brightness of the comparison surface had to be set at 1.5 ft.-L. and 5.8 ft.-L. These values were photometrically determined and differed slightly from the values which equated the illuminance of the standard

and comparison surfaces.[3] To equate the illuminance of the two surfaces (and the points of maximum brightness) the average brightness of the comparison surface had to be set at 1.32 ft.-L. and 5.31 ft.-L. These values were checked photometrically by measuring the brightnesses of the white sheets employed as backing.

Experiment III.—The standard and comparison surfaces used in Exp. III consisted of light and dark gray sheets of paper and are shown as B in Fig. 1. The arrangement of the surfaces and their backgrounds was similar to that of Exp. II. The area of the standard surface visible to *O* was $8\frac{1}{4}$ in. × $6\frac{1}{4}$ in. and of the comparison surface $10\frac{1}{4}$ in. × $5\frac{1}{4}$ in. Only the target surfaces and their backgrounds were again visible.

The reflectance of the standard surface was 44% and of the comparison surface 18%. The brightness of the standard surface was set at 2.4 ft.-L. and 5.5 ft.-L. The brightness of the background behind the standard surface for the two illumination levels was 2.31 ft.-L. and 5.27 ft.-L. To equate the brightnesses of the backgrounds, the brightness of the comparison surface had to be set at 1.19 ft.-L. and 2.68 ft.-L. These values were photometrically determined and differed from the calculated values which equate illuminance of the two surfaces .98 ft.-L. and 2.25 ft.-L.

Experiment IV.—The standard and comparison surfaces were medium gray and dark gray flannel cloths and are shown as C in Fig. 1. The surfaces were placed on a backboard 107 in. from *O* and set so that all that was visible to *O* were areas approximately 9 in. × 6 in. on the two surfaces. The reflectances of the standard and comparison surfaces were 15% and 4.5%. The average brightnesses of the standard surface at the lower and higher levels of illumination were .29 ft.-L. and 1.1 ft.-L. The maximum average brightness of the comparison surface corresponding to a switch setting of 40 was .96 ft.-L. Thus, at the higher illumination level it was not possible for *O* to equate the average brightnesses of the standard and comparison surfaces. This was a serious restriction since E was unable to set the illumination on the comparison surface for many *O*s at a level which was clearly above the illumination of the standard. Nine of the 15 *O*s set the comparison switch one or more times at position 40 when the higher illumination level was used. To equate the illuminances of the standard and comparison surfaces, *O* had to set the average brightness of the comparison surface to .09 ft.-L. and .33 ft.-L.

Experiment V.—The experimental setup was the same as that of Exp. IV except that metallic sparkles and sequins were distributed over the standard and comparison surfaces to produce highlights. As in Exp. IV, *E* was unable for many *O*s to set the illumination on the comparison surface clearly above that of the standard. Five *O*s set the comparison switch one or more times at the maximum position when the higher illumination level was used.

Experiment VI.—The standard and comparison surfaces each consisted of two pieces of wood joined together at right angles with one wood piece overlapping the other. The end side of the overlapping wood piece was visible to *O*, and on the comparison surface was painted black as shown in Fig. 1. The surfaces were placed so that the joined edge of each surface was centered facing *O* with the right and left sides extending backward in the form of a V. The distances from the joined edges of the two surfaces to *O* were 111 in. Twelve inches behind the standard and comparison surfaces were placed a gray and white background surface, respectively. The surfaces and backgrounds were arranged so that all that was visible to *O* were a 10 in. × 8 in. section of each side of the V-shaped surfaces, and the backgrounds above and to the right of the surfaces. The average reflectances of the standard and comparison surfaces were 40% and 6.4%. The average brightnesses of the standard surface at the lower and higher levels of illumination were 1 ft.-L. and 4 ft.-L. The brightnesses of the background behind the standard surface were .4 ft.-L. and 1.6 ft.-L. at the two illumination levels. The maximum brightness of the comparison surface was 1.52 ft.-L. At the higher illumination level, it was thus not possible for *O* to equate the average brightnesses of the standard and comparison surfaces. Unlike Exp. IV and V, the restriction in this case was not limiting since all *O*s set the switch considerably below the maximum setting. It was always possible for *E* to set the illuminance on the comparison surface at a level clearly too high. To equate the illuminances of the standard and comparison surfaces, *O* had to set the average brightness of the comparison surface to .13 ft.-L. and .64 ft.-L. No lightness matches were obtained of the target surfaces since the wooden surfaces could not be matched to the grays of the Hering scale.

[3] The difference is due to the fact that the light boxes were made up of individual bulbs which did not equally illuminate the target surface and background.

TABLE 1

MEDIANS OF *Os*' ILLUMINATION MATCHES

Exp.	Av. Brightness of Standard (Ft.-L.)	Med. Illumination Matches (Ft.-L.)	Confidence Intervals (Ft.-L.)
I	2.7	2.82	2.43–3.00
	6.8	6.40	6.10–7.73
II	2.4	2.00	1.50–2.10
	9.3	6.30	4.60–7.40
III	2.4	1.35	1.00–1.70
	5.5	2.75	2.10–3.70
IV	.29	.31	.25–.43
	1.1	.72	.51–.93
V	.29	.14	.1–.3
	1.1	.59	.33–.81
VI	1	.17	.12–.2
	4	.5	.24–.72
VII[a]	1	.22	.13–.29
	4	.46	.23–.78

[a] $N = 12$. A 96.1% confidence interval was formed by taking the third and tenth of the ordered medians.

Experiment VII.—Twelve of the *Os* employed in Exp. VI were rerun a week later with exactly the same arrangement as in Exp. VI except that the backgrounds were removed and only the target surfaces were visible.

RESULTS

Table 1 shows the sample median and a 96.5% confidence interval for each experiment and illumination level. The confidence interval was obtained by taking the fourth and twelfth of individual *Os*' medians arranged in order of magnitude (Dixon & Massey, 1957, p. 462) and gives the probability that the true median of the sample is contained within the interval formed. The intervals may be used to distinguish in an approximate way the relative influence of the different luminance variables on *Os*' matches.

Inspection of the table indicates that in Exp. I and IV, *Os* were unable to accurately judge surface illuminance. The *Os*' illumination matches for both levels of illumination of Exp. I and for the lower level of illumination of Exp. IV fall closer to the settings which equate the average brightnesses of the standard and comparison surfaces. The sample confidence intervals in all three cases contain the settings equating the average brightnesses of the surfaces but not those which equate the points of maximum brightness or illuminances of the surfaces. The *Os*' illumination matches at the higher level of illumination of Exp. IV fall between the settings which equate the average brightness and illuminance of the standard and comparison surfaces. They, however, have been displaced downward to an unknown extent by the failure of the light box illuminating the comparison surface to go high enough and their meaning is, therefore, unclear.

The presence of white backgrounds behind the standard and comparison surfaces in Exp. II and III changed *Os*' illumination matches in the direction of accurately judging surface illuminance. Unlike Exp. I and IV, the sample confidence intervals now fall below the settings which equate the average brightnesses of the standard and comparison surfaces and contain (except at the lower illumination levels of Exp. II and III) the settings which equate the illuminnances of the surfaces. That this change is due to *Os*' matches being strongly influenced by the brightnesses of the backgrounds is indicated in that the sample confidence intervals contain the settings which equate the background brightnesses. The introduction of metallic highlights in Exp. V also shifted *Os*' illumination matches away from an equating of the average brightnesses obtained in Exp. IV. A Mann-Whitney *U* test of the lowering of illumination matches obtained in Exp. V showed that the *P* values for the lower and higher

TABLE 2

MEDIANS OF *Os*' LIGHTNESS MATCHES OF STANDARD AND COMPARISON SURFACES AND THEIR BACKGROUNDS

Exp.	Av. Brightness of Standard (ft-L)	Standard		Comparison	
		Surface	Background	Surface	Background
I	2.7 6.8	43% 43%		25% 32%	
II	2.4 9.3	43% 43%	63% 63%	32% 32%	63% 63%
III	2.4 5.5	63% 63%	63% 63%	25% 25%	63% 63%
IV	.29 1.1	37% 43%		37% 43%	
V	.29 1.1	16% 19%		9% 9%	
VI	1 4		32% 43%		63% 63%

Note.—No lightness matches were obtained of the target surfaces in Exp. VI and VII because the wooden targets could not be matched to the grays of the Hering scale.

illumination levels were $<.02$ and .1, respectively. The range of *Os*' judgments as indicated by the confidence intervals, however, shows that the metallic highlights were not as effective in shifting *Os*' illumination matches as introducing a white background.

Table 1 shows that in Exp. VI and VII, *Os* tended to judge surface illuminance accurately. In both experiments, the settings equating the surface illuminances are contained in the sample confidence intervals. The experiments do not permit an analysis of the variables of luminance reflectance which might govern *Os*' matches. Though not indicated by the median illumination matches, the removal of the background surfaces raised *Os*' illumination matches slightly in Exp. VII (Sign test $P < .05$ for both the lower and higher illumination levels).

Table 2 shows the medians of *Os*' lightness matches of the standard and comparison surfaces and their backgrounds for the different experiments. Inspection of the table indicates that except for Exp. IV lightness constancy was good, with *Os*' matches clearly distinguishing between the lightnesses of the standard and comparison surfaces. Sign tests of the reflectances of the grays on the Hering chart which *Os* matched as equal in lightness to the standard and comparison surfaces were all significant at the .001 level. In Exp. IV, lightness constancy was impaired, and the medians of *Os*' lightness matches of the standard and comparison surfaces were identical. Sign tests of the differences in the lightness matches obtained for the standard and comparison surfaces showed P for both levels of illumination to be greater than .5. Tests of the differences in lightness matches occurring with increasing illuminance in Exp. I, IV, and V were not significant ($P > .1$). The increase in the lightness matches of the background surface occurring in Exp. VI was significant at the .01 level. Though explicit lightness matches were not obtained in Exp. VI and VII, *Os*' verbal reports

indicated a high degree of constancy of the standard and comparison surfaces.

A comparison of Tables 1 and 2 shows that the relationship between judgments of surface lightness and illuminance occurring in the various experiments was complex and not even notably consistent. In Exp. II, III, VI, and VII, lightness constancy was accompanied by a relatively accurate judgment of surface illuminance. However, in Exp. I, lightness constancy was accompanied by an inaccurate judgment of surface illuminance while in Exp. IV, the same kind of inaccurate judgment of surface illuminance was accompanied by impaired constancy. The introduction of metallic highlights in Exp. V produced lightness constancy but failed to yield corresponding accurate judgments of surface illuminance.

Discussion

The results of the experiments need to be examined with respect to (*a*) the basis for the judgment of illumination and (*b*) the relationship between judgments of surface lightness and illumination.

When the light source is not visible, an individual is able to judge the illumination of a surface only by means of the reflected luminance. There are, however, many variables of the reflected luminance which vary with the illumination and can serve as indications of it. Changes of illumination may be indicated, for example, by the distinctness of the surface texture, the average brightness of a surface, the points of maximum brightness of a surface, background brightnesses, the brightness of highlights, shadows, edge gradients and so forth. Observation of *Os*' behavior in the experiments suggested that in judging the surface illuminance *Os* developed a selective orientation to the stimulus variables available. Questioning of *Os* indicated that a few consciously chose a criterion (e.g., attempting to match the brightness of the metallic highlights in Exp. V) though most reported their judgments to be based on an overall visual impression which they were unable to analyze. Thus, it is likely that the degree to which different stimulus variables influence illumination judgments varies depending upon surface textures, finishes, spatial arrangements of neighboring surfaces, intensity levels and other situational factors. Individual differences may arise because *Os* respond differently to the stimulus variables present. A complete description of all possible variables and their degree of utilization is a sizeable task of little inherent interest. However, to answer some of the questions connected with the problems of surface perception, the perception of luminousness, and lightness constancy it is useful to be able to identify the principal variables.

The results of these experiments and those of Beck (1959) indicate the operation of three kinds of variables depending upon the stimulus situation. In the work reported by Beck (1959) with speckled and spotted surfaces and in Exp. II and III, the illumination judgments were strongly influenced by the intensity of the reflected light from clearly discriminable areas of high surface or background brightness. In comparison, in Exp. I and IV, where no clearly discriminable areas of high surface or background brightness were present (points of high brightness were part of the brightness gradients which provide the basis for the perception of a particular surface texture), the illumination judgments were strongly influenced by the average intensity reflected from the surface. In all these experiments, the surfaces were placed so that neither shadows nor highlights were present. In Exp. V, VI, and VII highlights were present as a result of the metallic sparkles and the wood finish. The accurate judgments of illumination in Exp. VI and VII suggest that *Os*' illumination judgments were based on the absolute or relative brightness of the highlights present on the wood surfaces. The failure of the metallic highlights in Exp. V to yield more accurate judgments of illumination may have been due to their intense brightness which made it difficult

to discriminate illumination changes. It should be noted that the variables which have been separately considered, namely, highlights, background brightnesses, and surface brightnesses ordinarily are present together and presumably combine to give an impression of the illumination. Consequently, care is needed in generalizing the results beyond the simplified conditions of the experiments.

According to the views that judgments of lightness and illumination are coupled psychologically, lightness constancy should fail badly where judgments of the incident illumination are close or equal to the intensity of the reflected light which enters the eye. In Exp. I, this was clearly not the case. Lightness constancy was accompanied by completely inaccurate judgments of the incident illumination. The results of the experiments suggest an alternative explanation of the connection of judgments of lightness and illumination based upon the relationships of the individual correlates. Constancy in Exp. I may be conjectured to be due to the presence of sharp contours within the surface. Wallach (1948) has shown that lightness is a function of the relationships of stimulus intensities in adjacent areas. The failure of constancy in Exp. IV would, thus, be due to the absence of sharp brightness gradients because of the diffuse reflectance of the flannel cloth. Judgments of illumination, however, in both Exp. I and IV would be inaccurate because they are at least in part a function of the average intensity of the light reflected. Support for the conjecture occurs in Exp. V where lightness constancy is obtained by the introduction of metallic highlights which also introduce sharp brightness gradients within the surface. The change of a disk from a dimly illuminated white to a strongly illuminated black with the introduction of a piece of white paper in Gelb's experiment may be similarly explained if judged lightness is taken to be a function of the relationships of stimulus intensities as indicted by the study of Stewart (1959) and judged illumination is taken to be a function of the maximum intensity as indicated by Exp. III.

Summary

Seven experiments studied the ways in which judgments of surface illumination and lightness varied with different surface textures and the presence of a background. Binocular matching using the method of adjustment was employed with the standard and comparison surfaces being viewed simultaneously, each with monocular vision and a motionless head, in a completely dark room. The experiments indicated judgments of illumination to be a function of variables of the reflected luminance (such as the surface and background brightnesses, the brightness of highlights) which depended upon the stimulus situation. The relationship between judgments of surface illumination and lightness was found not to be a unique coupling. An alternative analysis was proposed based upon the interrelations of the correlates for the judged lightness and illumination of a surface.

References

Beck, J. Stimulus correlates for the judged illumination of a surface. *J. exp. Psychol.*, 1959, **58**, 267–274.

Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1957.

Helmholtz, H. *Physiological optics.* (Trans. by J. C. P. Southall) Rochester: Optical Society of America, 1925.

Henneman, R. H. A photometric study of the perception of object color. *Arch. Psychol., NY*, 1935, No. 179.

Koffka, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

MacLeod, R. B. Brightness constancy in unrecognized shadows. *J. exp. Psychol.*, 1940, **27**, 1–22.

MacLeod, R. B. The effects of artificial penumbrae on the brightness of included areas. In, *Miscellanea Psychologica, Albert Michotte.* Paris: Libraire Philosophique, 1947.

Stewart, E. C. The Gelb effect. *J. exp. Psychol.*, 1959, **57**, 235–242.

Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.

Wallach, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received March 21, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 376-378

MATHEMATICAL SOLUTION IN THE ACQUISITION OF A VERBAL CR

J. P. DAS

Utkal University, Cuttack, India

Grant, Hake, and Hornseth (1951) have noticed a certain peculiarity in *S*'s behavior during the acquisition of a verbal CR. Humphreys' (1939) "guessing game" was used by them to establish verbal CRs to stochastically different stimulus rates. The stimulus was a lamp and *S*s were required to guess whether the lamp would or would not flash on. The light flashed 0, 25, 50, 75, and 100% of the time in separate experimental sessions. If *S*s behaved according to the mathematical solution for getting maximum hits, they should have consistently guessed that the lamp would not flash when the probabilities of flashing were lower than .50 and the opposite when the probabilities were higher than .50. Instead, *S*s adjusted their rate of emission to equal the rate of the flashing of the lamp. Estes (1954) has reported similar behavior and shows that *S*s learned that the rate of response can be accounted for by a probability learning model. Grant (1953) suggested that *S* might choose a mathematically optimal solution if he were continually reminded of his instructions to maximize the number of accurate "guesses."

The present study was carried out in an attempt to find a means of making *S*s adopt the mathematical solution. One of the most obvious means is to give *S* more information about his performance by allowing him to evaluate his performance from time to time. A somewhat modified form of the "guessing game" was used. Instead of flashing or not flashing lamps, a series of *yeses* and *noes* were spoken. The *S* was asked to predict whether *yes* or *no* would occur next. After *S* had made his response, either *yes* with a probability of .25 or *no* with a probability of .75 occurred according to a fixed random schedule independent of *S*'s response.

METHOD

A random series of 75 *yeses* and 25 *nocs* was tape recorded with a fixed interval of 10 sec. between any two words. The *S*s were required to guess during this interval whether the next word would be a *yes* or a *no*. The main experiment was preceded by a practice test comprising of 10 trials to familiarize *S*s with the procedure. Altogether 5 blocks of 25 trials were presented during the main experiment, the last block being a repetition of the first.

The *S*s were instructed to write down their guesses on the left-hand side of the record sheet provided and the corresponding "answers" as heard from the tape recorder on the right-hand side. It was insured that *S*s put down their guesses before the event occurred.

A total of 140 college students aged between 16 and 21 served as *S*s. They were tested in five groups under the following five conditions: (*a*) Condition NR (nonreinforced). The 23 *S*s were not given the benefit of the knowledge of their performance. After every 25 trials a 1-min. rest was allowed, after which they were asked to turn to a new page for the next block of 25 trials. (*b*) Condition R_{25}. The 28 *S*s were asked to count their number of hits at the end of every 25 trials, and then to read their scores aloud and turn to a new page for the next block of 25 trials. (*c*) Condition R_5. The 38 *S*s were asked to count the number of hits after every 5 trials and add these up at the end of every 25 trials, and asked to turn a new page to be ready to do the same for the next block of 25 trials. (*d*) Condition R_{no}. The 27 *S*s were told that only *no*-hits would earn

points. In all other respects their instructions were similar to Group R_5. (*e*) Condition $R_{no\text{-}yes}$. The 23 *Ss* were told that for each *no*-hit, they would gain a point, and for each *yes*-hit they would lose a point. So the total score would be the total number of *no* hits minus the total number of *yes* hits. In all other respects their instructions were similar to Group R_5.

It was hoped that the last group, especially, should tend to adopt the mathematical solution; that is, always giving a *no* guess.

Results and Discussion

Curves showing the gradual decrement of *yes* responses are shown in Fig. 1. The procedure for Group NR is comparable to the 25% procedure of Grant, Hake, and Hornseth (1951) and gives the same 25% performance in the fifth block of trials that the earlier investigators report.

Group R_{25} does worse with knowledge of the results of its performance, whereas, Group R_5 almost reaches 20% *yes* responses in the last block of trials.

The differences between the Groups R_5 and NR ($P < .05$) and between Groups R_5 and R_{25} ($P < .01$) were significant, as shown in Table 1. But the significant difference between Groups R_5 and NR only partially confirms our original conjecture. Group

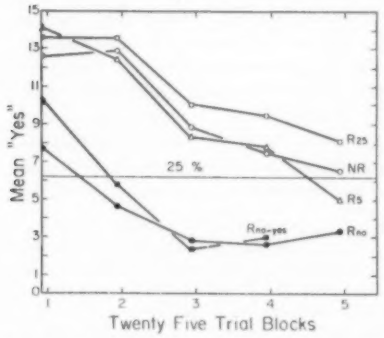


Fig. 1. Mean *yes* responses on successive 25-trial blocks for the five groups of *Ss*.

TABLE 1

Comparisons of Frequencies of *Yes* Responses in Selected Blocks of Trials

Group	Block	*N*	*Yes* Responses	
			Mean	*SD*
NR	5	23	6.74	2.50
R_{25}	5	28	8.28	3.23
R_5	5	38	5.13	2.68
R_{no}	4	27	2.74	2.92
$R_{no\text{-}yes}$	3	23	2.44	2.93

t (NR vs. R_{25}) = 1.89, *ns*.
t (NR vs. R_5) = 2.32, $P < .05$.
t (R_{25} vs. R_5) = 3.71, $P < .01$.
t (R_5 vs. R_{no}) = 3.46, $P < .01$.
t (R_{no} vs. $R_{no\text{-}yes}$) = .38, *ns*.

R_5 did not use the optimal solution even though the percentage of *yes* responses was reduced to nearly 20. But the last two groups fare better in this respect. Group R_{no} gives the lowest number of *yes* responses in the fourth block of trials, about 11% of the total number of responses. Group $R_{no\text{-}yes}$ reached its lowest number of *yes* responses still earlier, in the third block of Trials with less than 10% *yes* responses. The difference between Groups R_5 and R_{no} is significant at the .01 level; this is also true of the difference between Groups R_5 and $R_{no\text{-}yes}$.

The results so far discussed show that even the last group as a whole did not attain the mathematical solution. Since an *S* who attains the mathematical solution should consistently give zero *yes* responses, the number of *yes* responses of each *S* per 25 trial blocks was examined individually. It was found that none of the *Ss* of Groups NR and R_{25} gave zero *yes* response in the last block of trials. In the other groups the frequency of zero *yes* responses was as follows: 1 of 38 *Ss* in Group R_5 in Block 5; 8 of 27 *Ss* (30%) of Group

R_{no} in Block 4; and 11 of 23 *Ss* (48%) in Group$_{no\text{-}yes}$ in Block 3. Since Block 4 of Group R_{no} and Block 3 of Group$_{no\text{-}yes}$ had the lowest means, the zero *yes* scores in those blocks are reported. Other data of similar interest are the numbers of *Ss* who gave only one or two *yes* responses in the last block of trials: 2 in Group NR, 2 in Group R_{2s}, 4 in Group R_{s}, 5 in Group R_{no}, and 4 in Group$_{no\text{-}yes}$.

Some *Ss* of Groups NR, R_{2s}, and R_{s} verbalized the mathematical solution in making statements like "all guesses would be correct if one writes *no* all the time." Most *Ss* of Groups R_{no} and $R_{no\text{-}yes}$ did so. These statements were spontaneously made during the experiment or immediately afterwards. No doubt, everyone in the group heard these. In spite of this, it is surprising that none of the *Ss* in Groups NR and R_{2s}, and less than half of the *Ss* in Group $R_{no\text{-}yes}$ actually attained the mathematical solution. Some dissociation between the verbal and behavioral levels is shown by this.

Summary

The *Ss* in this experiment guessed whether *yes* or *no* would occur in 125 random occurrences in which the probability was .25 for *yes* and .75 for *no*. Several conditions involving increasing knowledge of results were imposed to encourage *Ss* (124 in 5 groups) to adopt a games solution by predicting the more frequently occurring event on all trials. Although the number of zero and near-zero *yes* responses increased progressively as *Ss* were given more knowledge of results, even *Ss* in the group given the most knowledge—told that the score was the number of *no* hits minus the number of *yes* hits—attained the games solution less than 50% of the time.

References

Estes, W. K. Individual behavior in uncertain situations: An interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954.

Grant, D. A. Information theory and the discrimination of sequences in stimulus events. In, *Current trends in information theory*. Pittsburgh: Univer. Pittsburgh Press, 1953.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

(Received March 25, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 379–388

THE DISCRIMINATION OF RELATIVE ONSET-TIME OF THE COMPONENTS OF CERTAIN SPEECH AND NONSPEECH PATTERNS[1]

A. M. LIBERMAN, KATHERINE S. HARRIS, JO ANN KINNEY,[2] AND H. LANE[3]

Haskins Laboratories and University of Connecticut

In earlier experiments (Griffith, 1958; Liberman, Harris, Hoffman, & Griffith, 1957) with synthetic stop consonants it was found that discrimination of a constant acoustic difference was more acute across phoneme boundaries (between /b/ and /d/ and between /d/ and /g/) than in the middle of the phoneme categories. Indeed, the discrimination was relatively so good across the boundaries and so poor within the categories as to suggest that *S*s could only respond to these sounds categorically (i.e., as phonemes) and could hear no other differences among them. It was possible to test this suggestion quantitatively by determining whether the discriminability of the sounds was predictable from the frequency with which they had been identified as belonging in one or another phoneme class. In general, the locations of the peaks and valleys of the discrimination functions were well accounted for on this basis. The level of discrimination, on the other hand, was on the whole somewhat higher than predicted, but not by a very great amount. Thus, differential sensitivity was in this case only slightly better than absolute identification. Such a result must appear unusual when it is remembered that in the typical psychophysical experiment the number of discriminably different stimuli is found to be many times greater than the number which can be labeled or identified absolutely (see Chapanis, 1956; Garner, 1953; Miller, 1956; Pollack, 1952, 1953).

The outcome of the experiment with the synthetic stop consonants can be interpreted in either of two ways. On the one hand it is possible that the peaks and valleys in the discrimination function are given innately. Alternatively, it can be assumed that two sounds which a listener must normally respond to in the same way come, by some learning process, to be less discriminable than two equally different sounds to which he must typically attach different phoneme labels.

If it is assumed that the discrimination functions have been radically modified by learning, there remains a further question concerning the direction of the effect. One possibility is that the innately given discrimination was originally as poor as that obtained for the least discriminable pair of stimuli, and that discriminations in the vicinity of phoneme boundaries have been sharpened by something like the process of acquired distinctiveness. The contrary possibility is that in its virgin state discrimination was everywhere as good as that at the peaks of the functions, and that long experience in responding similarly to stimuli

[1] This research was supported by the Operational Applications Laboratory of the Air Force Cambridge Research Center in connection with Contract AF 19(604)-2285.

[2] Now at Medical Research Laboratory, United States Naval Submarine Base, New London, Connecticut.

[3] Now at the University of Michigan.

within the same phoneme category has, through acquired similarity, significantly reduced the discriminability of those sounds. It is also possible, of course, that the original discrimination lay between the two extremes. In that case the results would be taken to represent the operation of both processes: acquired distinctiveness near phoneme boundaries and acquired similarity within the phoneme category.

The experiment to be reported was designed to investigate the relation between phoneme labeling and discrimination in the case of an acoustic and linguistic distinction different from the ones studied earlier, and, hopefully, to throw some light on the issues discussed above. The phonemes to be used are the voiced and voiceless alveolar stops /d, t/.

One reason for investigating /d,t/ is that Liberman, Delattre, and Cooper (1958) have shown that many *S*s can consistently identify as /d/ or /t/ synthetic patterns that differ by only small amounts of the appropriate acoustic cue. We may, then, expect to find with this phonemic distinction, as we have earlier found with /b,d,g/ (Liberman et al., 1957; Griffith, 1958), that the discrimination function has a sharp peak at the phoneme boundary.

A second reason for selecting /d,t/ is that the discrimination of these phonemes can be made to depend on a very simple acoustic variable—the relative displacement in time of onset between the first and second formants.[4] It is possible to study discrimination of this simple time cue in nonspeech contexts, thus providing a baseline control in terms of which one might infer whether the effects obtained with the speech sounds are the results of learning, and, if so, whether they represent acquired distinctiveness, acquired similarity, or both. A variety of different nonspeech controls is, in this case, possible and necessary. One such control has been undertaken with this experiment.

[4] A "formant" is a relatively high concentration of acoustic energy within a restricted region of the frequency spectrum. The formants are numbered according to their positions on the frequency scale, the lowest formant being called the "first," the next higher the "second," and so on.

Method

Apparatus.—All the stimuli of this experiment were generated by using the Pattern Playback to convert hand-painted spectrograms into sound. The Pattern Playback and the techniques for using it are described in earlier papers (Cooper, 1950, 1953; Cooper, Liberman, & Borst, 1951).

Stimuli.—The spectrograms used to produce the stimuli are shown in Fig. 1. In the top row are patterns that are heard by most *S*s as /do/ or /to/ when converted into sound by the Pattern Playback. These will be referred to as the "speech" stimuli. As the figure shows, the stimulus variable which distinguishes these patterns is the delay in the onset of the first formant relative to the second and third formants. (This variable had been found in an earlier study by Liberman et al. (1958) to be sufficient for the perceived distinction between /d/ and /t/.) The delays vary in 10-msec. steps, from 0-msec. delay, as in the pattern at the extreme left, to 60-msec. delay, as shown at the right. The other characteristics of the pattern are those appropriate to an alveolar stop consonant in initial position before the vowel /o/. For convenience, these stimuli will be referred to by number, from 0 through 60, as indicated in Fig. 1.

One of the purposes of this study was to compare the discriminability of the synthetic speech sounds described above with those of a control set which were as much like them as possible, and yet not perceived as speech. The stimuli that were used for this purpose are shown in the bottom row of Fig. 1. These patterns, which will be referred to as the "control" stimuli, were made simply by turning the speech spectrograms upside down before running them through the Pattern Playback. The stimuli that are produced by this maneuver differ from each other in xactly the same way that the speech stimuli

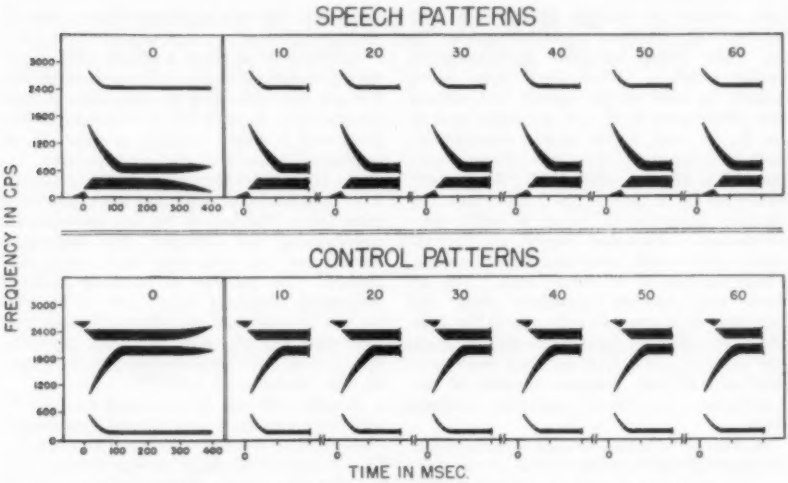


FIG. 1. Spectrographic patterns which were converted to sound by the Pattern Playback to form the stimuli of the experiment.

differ—namely, by the relative time of onset of the formants—and, like the speech stimuli, they can be discriminated only on that basis. As can be seen from an inspection of the control patterns, one rather small change was made in addition to the frequency inversion. The relatively "thin" formant, which is at the top of the pattern (Formant 3) in the speech stimuli and at the bottom (Formant 1) in the control, was altered in the control so as to have a falling transition in place of the rising transition that would have resulted from the simple inversion. This was to provide a further precaution against the possibility that the control stimuli would resemble speech. An additional difference between the speech and control stimuli is in the relative intensities of the three formants. In the speech stimuli the first and second formants were of essentially equal intensity, while the thin third formant was approximately 5 db. less intense. In the control stimuli the first and second formants have become the second and third and are, again, of equal intensity; however, the relatively thin formant, which has become the first formant, is now only 3 db. below the other two.

Careful listening to the control stimuli revealed that they did not sound at all like speech. Certainly, no phonemic distinction was conveyed by the stimulus variable.

All the patterns of Fig. 1 were converted into sound and recorded on magnetic tape. By copying, cutting, and splicing the magnetic tape, the various arrangements described below were prepared.

Stimulus presentation.—The purpose of this experiment made it necessary that we measure the discriminability of both the speech and control stimuli, and that we find the location and shape of the phoneme boundary. Discriminability of the speech and control stimuli was measured by the ABX technique. The *Ss* were told that A and B would always be different, and that X would be identical with the one or the other. They were asked to judge, on whatever basis they could, whether X was like A or like B, and to guess if necessary.

The speech and control stimuli were grouped separately for ABX presentation, being arranged in magnetic tape sections that consisted of speech stimuli or control stimuli, but never both in the same tape section. The procedures for making the speech tapes and the control tapes were identical and are described below. This description has been written as if it were for only one of the two sets of stimuli (speech or control).

The various triads were made by pairing each stimulus with another stimulus having an onset delay that differed in the amount of 10, 20, or 30 msec. These pairs formed 1-, 2-, and 3-step intervals, respectively. Thus, for the 1-step intervals, Stimulus 0 was paired with Stimulus 10, Stimulus 10 with Stimulus 20, etc. The 2-step intervals

were formed by pairing Stimulus 0 with Stimulus 20, Stimulus 10 with Stimulus 30, etc. The 3-step intervals were made in similar fashion. Since there were seven stimuli (in each of the speech and control sets), there were 6, 5, and 4 comparisons in the 1-, 2-, and 3-step series, respectively. Each stimulus comparison was arranged into four ABX (ABA, ABB, BAB, BAA) permutations for a total of 60 ABX triads (15 stimulus comparisons times 4 ABX permutations). Recorded copies of the 60 triads were made and distributed among 8 tape sections of 15 triads each. The 15 triads were ordered randomly with the restriction that one and only one of the four ABX permutations be represented in each tape section. The tape sections were prepared with 0.5 sec. between members of the stimulus triad and 4.0-sec. separation between triads.

In order to ascertain for each *S* the position and shape of the boundary between the "d" and "t" phonemes, each *S* was presented the same speech tapes that were used to measure discrimination, but in this case *S* was asked to identify each member of the triad as /d/ or /t/ and to guess if necessary. This aspect of the procedure will be referred to as "phoneme labeling." [5]

Subjects.—The *Ss* were summer students at Columbia University. All were paid volunteers, and had no knowledge of the purpose of the experiment.

In order to determine whether discrimination is relatively sharper at the phoneme boundary, it is, of course, necessary that *S* assign the phoneme labels with a high degree of consistency, and that the region of uncertainty between /d/ and /t/ be narrow. We therefore preselected our *Ss*. Twenty *Ss* were each presented four speech tapes with phoneme labeling instructions. The 11 *Ss* whose judgments provided the sharpest phoneme boundaries were selected for the experiment. The difference between these 11 *Ss* and the 9 who were rejected was not very great. The judgments obtained during the selection session were not included in the data of the experiment.

The 11 *Ss* served subsequently for approximately 30 min. per day, 5 days a week, for a total of 20 sessions. In Sessions 1 and 2 only labeling tapes were presented. Speech and control ABX tapes were introduced in Session 3. In each of the remaining sessions, *Ss* heard at least one tape of each type.

In these 20 sessions, 4 speech ABX tapes were presented 6 times, and 4 were presented 5 times, thus providing 44 judgments of each comparison. Each of the 8 control tapes was presented 5 times, making a total of 40 judgments for each stimulus comparison.

Six of the tapes were presented for phoneme labeling four times, while two were presented three times. Because of the way the stimulus comparisons are arranged, the individual stimuli are not presented with equal frequency. The number of phoneme labeling judgments obtained from each *S* for the various stimuli was as follows: Stimulus 0, 135; Stimulus 10, 180; Stimulus 20, 225; Stimulus 30, 270; Stimulus 40, 225; Stimulus 50, 180; Stimulus 60, 135.

Results will also be presented for two *Ss* who were run in a preliminary experiment conducted at the University of Connecticut. The procedure for these *Ss* was the same as for the larger group, except that the number of ABX judgments was approximately 50% greater and the number of phoneme labeling judgments about 50% less than with the larger group.

Results

In Fig. 2 are plotted the responses that were obtained when the triads of synthetic speech stimuli were presented for phoneme labeling, i.e., for identification as /d/ or /t/. Each function represents the data for one *S*, and is based on responses to all presentations of the stimuli. The values on the abscissa, which represent amount of first-formant cutback in milliseconds, are progressively displaced (by 10 msec. of cutback) for each function in order that all the data may be shown on a single graph. A reference for the displacement of the abscissa is provided by the circled points, which indicate the percentage of /do/ responses for Stimulus 20.

The phoneme labeling curves are quite similar, both in position and in form. All the functions but one pass through the 50% point somewhere between Stimulus 20 and Stimulus 30;

[5] In the earlier study of /b,d,g/ (Liberman et al., 1957) the speech stimuli were presented separately for phoneme labeling, and not in triads as in the present experiment.

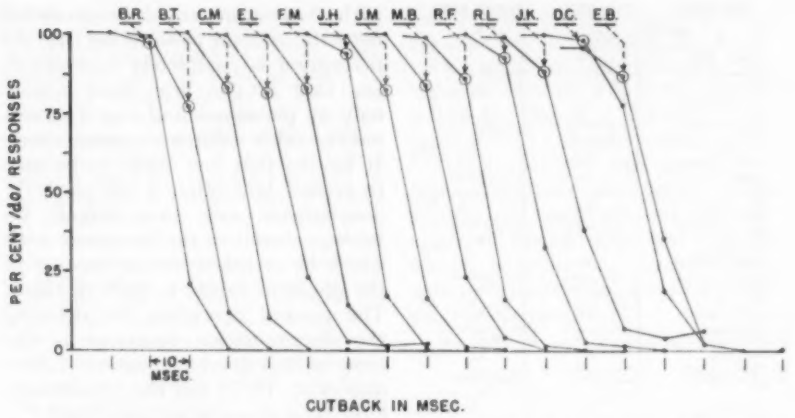


FIG. 2. Response functions for each of the 13 *S*s, showing how phoneme labels were assigned to the stimuli. The circled points represent, for each *S*, the percentage of /do/ responses to the stimulus that had 20 msec. of cutback. So that the functions might be better separated for the eye, each one has been started just beyond the point that would represent the results at 0 msec. of cutback. The /do/ responses at that point (100% for 10 *S*s, 99% for 2 *S*s, and 96% for 1 *S*) can, however, be inferred from the graphs.

the remaining function crosses the midpoint between Stimulus 30 and Stimulus 40. For all *S*s the phoneme boundary between /d/ and /t/ is quite sharp; in every case a change of 10 msec. in the first-formant cutback is sufficient to shift the responses from 75% /d/ to 75% /t/, and there are no large inversions in the responses.

One purpose of this experiment was to determine whether or not discrimination is better at the phoneme boundary than in the middle of the phoneme category. This question is answered by the graphs for individual *S*s in Fig. 3. These are based on data that were obtained when the stimuli were presented in ABX triads and *S*s were asked to judge whether X was identical with A or with B. The data of Subject BR may be used to illustrate the results. The labeling data of Fig. 2 show that for this *S* the /d/ responses shifted from 97% at Stimulus 20 to 18% at Stimulus 30, so that the phoneme boundary may be said to lie between these two stimuli. The upper left-hand graph of Fig. 3 (solid lines) shows the discrimination results for this *S* when the stimuli being compared differed by one step (10 msec.) on the stimulus scale. It can be seen that there is a peak in the discrimination function at the point which represents the comparison of Stimulus 20 and Stimulus 30, the two stimuli which straddle the phoneme boundary. In the 2-step case Stimulus 10 and Stimulus 30 fall on opposite sides of the phoneme boundary, as do Stimulus 20 and Stimulus 40, and it can be seen that discrimination is, indeed, better in these cases than in others. For the 3-step situation all stimulus pairs except Stimulus 30 and Stimulus 60 cross the phoneme boundary, and discrimination is at a high level; the 3-step pair in which the stimuli (30 and 60) lie on the same side of the boundary is seen to be considerably less discriminable.

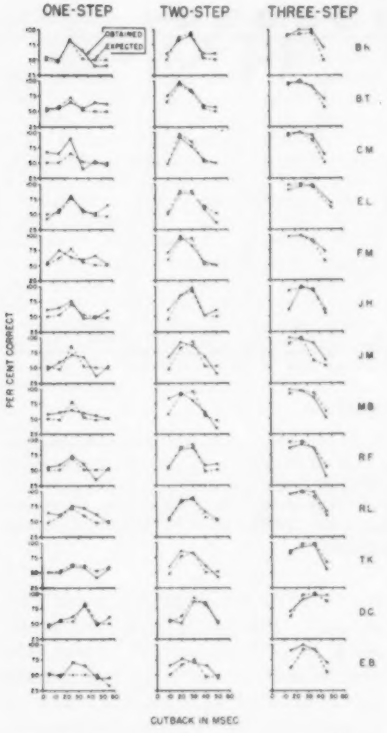


FIG. 3. Expected and obtained discrimination functions for the 1-, 2-, and 3-step differences among the synthetic speech stimuli.

The effects observed in the case of BR were obtained to a very large extent with all *S*s. In general, discrimination was better for a given stimulus difference when the two stimuli lay on opposite sides of a phoneme boundary than when the stimuli were within the same phoneme category. This effect is very apparent for the 2- and 3-step data, though perhaps not quite so obvious for the 1-step comparison.

In order to quantify the effect just described it is useful to compare the discrimination functions of Fig. 3 with those that would be predicted from the extreme assumption that *S*'s perception is completely categorical, i.e., that he can hear these sounds only as phonemes and can discriminate no other differences among them. It follows that one ought to be able to predict how often *S* will correctly discriminate any two stimuli by taking account of the frequency with which he attached one or another of the phoneme labels to each of them. The present procedure for deriving the discrimination functions is the same as that developed earlier (Liberman et al., 1957), but the calculations have been changed and simplified.[6]

An example will make the method of calculation clear. Suppose that the presented triad of stimuli is Stimulus 0–Stimulus 20–Stimulus 0. When these stimuli are presented with phoneme labeling instructions, *S* is asked to identify each one as /d/ or /t/. Neglecting for the moment how *S*s actually labeled these stimuli and considering only how they might have done so, we see that there are eight possible sequences of phoneme labels for any triad. For the purpose of predicting how *S* would respond in the ABX test of discrimination, these eight sequences can be divided into three classes. One class, which consists of two sequences (/d/ /t/ /d/ and /t/ /d/ /t/), would lead S to the correct response, "A." A second class, consisting of two sequences (/d/ /t/ /t/ and /t/ /d/ /d/), would lead *S* to the incorrect response, "B." The remaining four sequences (/t/ /t/ /t/, /d/ /d/ /d/, /d/ /d/ /t/, and /t/ /t/ /d/) would leave *S* no alternative but to guess, and he would, then, presumably respond "A" half the time and "B" half the time (probability of being correct = .5). Now suppose the triad (Stimulus 0–Stimulus 20–Stimulus 0) is presented to *S* four times for phoneme

[6] The simpler method can be used when the stimuli are presented in the same triads for labeling and discrimination. It yields the same result as the earlier method provided the various triad contexts have had no effect on *S*s' assignment of phoneme labels. For the data of several *S*s, discrimination functions were calculated according to the earlier method and were found to be almost identical with those produced by the simpler method.

labeling, and suppose, further, that he responds three times with the sequence /d/ /d/ /d/ and once with /d/ /t/ /d/. The sequence /d/ /d/ /d/ would produce a .5 probability of being correct in the ABX discrimination; the sequence /d/ /t/ /d/ would produce a probability of 1.0. Multiplying the probability of .5 by the three sequences which are calculated to yield it, and adding this to a probability of 1.0 times the one sequence that would yield it, gives 2.5. Dividing this quantity by four, which is the number of presentations, one obtains .63 as the average predicted percent correct in the ABX discrimination of Stimulus 0 and Stimulus 20. Since the comparison of Stimulus 0 and Stimulus 20 will be made not only in the ABA order of the example taken here, but also the orders ABB, BAB, and BAA, the predicted percentage correct for the discrimination of these two stimuli is the average of the predictions obtained from all four ABX permutations.

The discrimination functions that are predicted from the assumption of completely categorical perception are shown in Fig. 3 by the dotted lines. A comparison with the obtained functions shows that on the whole the fit is a good one. For the 1- and 2-step data there appears to be no systematic displacement of the predicted and obtained curves; the differences occur equally in both directions. In the 3-step case the average of the obtained discriminations is higher on the whole than predicted, but the difference is very small.

To the extent that the obtained discrimination data fit the predicted, it can be said that the peaks in the obtained functions do, indeed, lie at the phoneme boundaries, a conclusion that otherwise could have been tested only roughly and by point-for-point comparison of the discrimination and labeling data of Fig. 2. Moreover, and more generally, the close fit supports the conclusion that in perceiving these speech sounds the listeners were able to extract no information beyond that which was given by the phoneme categories.

One asks about discrimination functions like those of Fig. 3 whether they reflect the effects of learning on perception. The discriminability of the "control" stimuli was measured to provide evidence on this point. These stimuli contained essentially the same stimulus variations as the synthetic speech patterns, but were otherwise sufficiently altered as not to be clearly perceptible as speech.

In Fig. 4 are the discrimination

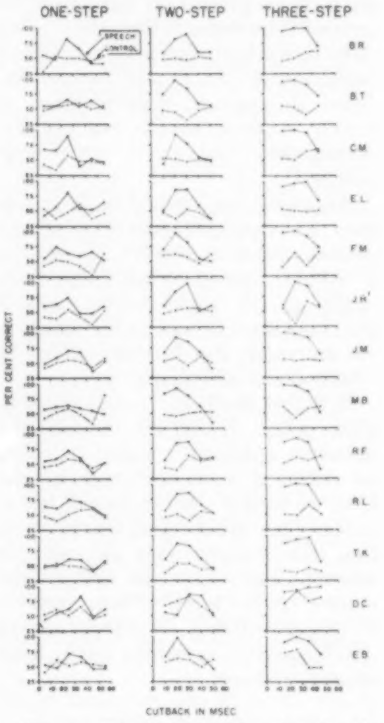


FIG. 4. Discrimination functions for the 1-, 2-, and 3-step differences among the nonspeech control stimuli. The discrimination functions that were obtained with the synthetic speech stimuli (and shown in Fig. 3 as "obtained" data) are reproduced here to facilitate comparison.

data obtained with the control stimuli. For purposes of easy comparison, the data obtained with the synthetic speech (and already shown in Fig. 3) are reproduced. Looking first at the control functions alone, one sees that there are no large or regularly occurring peaks at stimulus values corresponding to the position of the phoneme boundary or at any other points on the stimulus continuum. Comparison of the speech and control functions shows that the peaks of the speech discrimination functions for all *S*s tend in general to lie above the control functions. For the 1-step data this is not obviously true for all cases, but for the 2- and 3-step functions the speech discrimination is clearly superior to the control discrimination, and by a very wide margin.

Assuming the control to be a fair one, we conclude that the speech discrimination data reflect the effects of a large amount of learning. Inasmuch as the peaks of the speech discrimination lie above the control, we interpret the learning effect as acquired distinctiveness.

Acquired similarity would presumably occur, if at all, within the phoneme categories, where it would be reflected as a reduction in the level of speech discrimination below that of the control. One sees in Fig. 4 that the control data do not, in general, rise very far above the chance (50%) level. Thus, there is in this experiment no margin within which acquired similarity might have become manifest.

Discussion

To determine whether the increased discrimination at the phoneme boundary is an effect of learning on perception, and, if so, whether it represents acquired distinctiveness, acquired similarity, or both, we compared the discriminability of the speech stimuli with a set of nonspeech controls. Ideally, these control stimuli should have been identical with the speech stimuli, and yet not perceived as speech. In practice, such controls can only be approximated, since the acoustic pattern in which the relevant physical variable is imbedded must necessarily be somewhat different if, in the one case, the stimulus is to be perceived as a speech sound and in the other not. The control stimuli used in this experiment represent only an approximation to the ideal. They are not the only possible approximation, nor are they necessarily the best.

It can be counted an advantage of the control stimuli used here that the physical differences which cued the distinction between /do/ and /to/ in the speech stimuli (i.e., the relative difference in time of onset of first and second formants) were in one sense perfectly preserved, and that they constituted in the control, as they had in the speech series, the only acoustic differences among the stimuli. The control stimuli were appropriate, too, in that many of the constant acoustic features of the speech stimuli, such as formant width, the complex nature of formant onset, etc., were closely duplicated. Many of these features of the rather complicated stimulus pattern might well affect the discrimination of onset time, and make it quite different from what it would be with the most nearly comparable simple stimuli.

In several other respects, however, the control stimuli were less than perfect. First, it should be noted that the frequency level of the middle formant was different in the control stimuli from what it was in the speech. More significant, perhaps, is the fact that in the control stimuli the formant whose time of onset varied was at a higher frequency than the other two formants, while in the speech stimuli it lay at a lower frequency than the other formants. This may be important since masking effects are greater from low frequencies to high, and it is possible, therefore,

that the variations in formant onset were to some extent masked out in the control.

The point was made above that since the stimulus variations to be discriminated (the differences in onset-time) were imbedded in a complex pattern in the speech stimuli, it was appropriate that they be similarly imbedded in the control. It is nevertheless pertinent to inquire about the discriminability of essentially the same stimulus variable in a much simpler context—where the discrimination data might be considered to provide a psychophysical baseline—and to compare the results of such a study with those obtained here. A recent experiment by Hirsh (1959), in which *Ss* judged which of two pure tones (of different frequency) started first, is relevant to this point.

One finds in Hirsh's results no indication of the sharp peaks so clearly evident in the discrimination functions of the present experiment. It is also of interest that the discrimination of onset time was better in the speech patterns than in Hirsh's stimuli. For speech stimuli that straddle phoneme boundaries (those that were predicted from the phoneme labeling results to yield the best discrimination), the difference in time of onset required to give 75% correct judgments was slightly less than 12 msec. The 75% point in Hirsh's experiment was about 17 msec. This difference appears the more significant when one takes into account that the stimulus variable was in a much simpler acoustic context in Hirsh's experiment, and that his *Ss* were permitted to hear each stimulus pair as many times as they wished before making a judgment, while our *Ss* had to respond after each presentation. These comparisons support the assumption that the discrimination peak at the phoneme boundary is a result of learning.

If we put aside the results of the control, we are left with speech discrimination data that fit very closely a set of curves derived on the extreme assumption that *Ss*' discriminations were completely determined and limited by the phoneme labels they attached to the sounds. Such a result contrasts with the usual psychophysical finding, which is that many more stimuli can be discriminated than can be identified or labeled absolutely. It is difficult to conceive that discrimination and labeling would be equivalent, as they proved to be in this experiment, except on the assumption that the discrimination processes have been radically altered by long practice in attaching phoneme labels.[7]

Summary

An experiment was designed (*a*) to measure the discriminability of certain acoustic differences when they are cues for the perceived distinction between phonemes and when they lie entirely within a single phoneme category, and (*b*) to compare such discrimination data with those obtained when essentially the same acoustic differences occur in sounds that are not perceived as speech.

The speech sounds were synthetic approximations to the syllables /do/ and /to/, which differed only in the relative time of onset of first and second formants, this feature having previously been found to be sufficient to cue the distinction between /d/ and /t/ phonemes. To find the location of the phoneme boundary on the scale of relative onset-times, the various stimuli were presented in several random orders, with instructions to identify each sound as /do/ or /to/. To measure discrimination, the stimuli were presented in ABX fashion, and *Ss* were instructed to use whatever differences or similarities they could hear as a basis for deciding whether the third stimulus (X) was identical with the first (A) or the second (B).

It was found that discrimination was considerably better across a phoneme boundary than in the middle of a phoneme category. Indeed, this effect was so marked as to justify the assumption that the perception of these stimuli is essentially categorical—that is, that *S* can hear no differences among the stimuli beyond those that are revealed by the phoneme labels he applies to them. On that extreme assumption it was possible to use the frequencies with which the various stimuli were labeled /d/ or /t/ as a basis for predicting very accurately how well the stimuli could be discriminated.

A set of control stimuli—patterns of

[7] For a discussion of the mechanism which possibly underlies these learning effects, see Liberman (1957), pp. 121–123.

sound differing from each other in essentially the same way that the speech stimuli differed —was produced by inverting the speech patterns on the frequency scale. The discrimination data obtained with these inverted controls revealed no increase in discriminability at stimulus values in the region corresponding to the location of the phoneme boundary, and, in general, the discriminability of the control stimuli was much poorer than that of the speech. To the extent that the inverted stimuli are an appropriate control, it may be concluded that the sharpening of discrimination at the phoneme boundary is an effect of learning and, more specifically, that it represents a considerable amount of acquired distinctiveness.

References

Chapanis, A. S., & Halsey, R. M. Absolute judgments of spectrum colors. *J. Psychol.*, 1956, **42,** 99–103.

Cooper, F. S. Spectrum analysis. *J. Acoust. Soc. Amer.*, 1950, **22,** 761–762.

Cooper, F. S. Some instrumental aids to research on speech. In, *Report of the fourth annual round table meeting on linguistics and language teaching.* Washington, D. C.: Institute of Languages and Linguistics, Georgetown University, 1953. Pp. 46–53.

Cooper, F. S., Liberman, A. M., & Borst, J. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Nat. Acad. Sci.*, 1951, **37,** 318–325.

Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46,** 373–380.

Griffith, B. C. A study of the relation between phoneme labeling and discriminability in the perception of synthetic stop consonants. Unpublished doctoral dissertation, University of Connecticut, 1958.

Hirsh, I. J. Auditory perception of temporal order. *J. Acoust. Soc. Amer.*, 1959, **31,** 759–767.

Liberman, A. M. Some results of research on speech perception. *J. Acoust. Soc. Amer.*, 1957, **29,** 117–123.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Lang. Speech*, 1958, **1,** 153–157.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 1957, **54,** 358–368.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63,** 81–97.

Pollack, I. The information of elementary auditory displays. Part I. *J. Acoust. Soc. Amer.*, 1952, **24,** 745–749.

Pollack, I. The information of elementary auditory displays. Part II. *J. Acoust. Soc. Amer.*, 1953, **25,** 765–769.

(Received March 28, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 389–399

THE RELATION OF APPARENT BRIGHTNESS TO THE THRESHOLD FOR DIFFERENCES IN LUMINANCE[1]

ERIC G. HEINEMANN

Vassar College

When two luminous stimuli are present on neighboring regions of the retina each has an effect on the apparent brightness of the other ("simultaneous brightness induction") and also on the sensitivity to luminance differences of the region stimulated by the other. Each of these effects has been investigated quite thoroughly. The first extensive study of simultaneous brightness induction was done by Hess and Pretori (1894). More recent investigations concerned with the same phenomenon include those of Schouten and Ornstein (1939), Fry and Alpern (1953), Diamond (1953), and Heinemann (1955), to cite a few of the major references. The effect of presenting a stimulus on one region of the retina upon the luminance difference sensitivity in another region has been the subject of a large number of investigations of which those done by Cobb (1916), Holladay (1926), and Stiles (1929) are representative.

Though much is known of the effects of inducing fields (or "glare sources") upon apparent brightness and upon sensitivity individually, the relation between these two effects is not well understood. This is illustrated, for example, by the fact that both decreases in the apparent brightness of a test field and decreases in luminance difference sensitivity resulting from the presence of an inducing field are frequently treated as measures of "inhibition." Actually, these cannot both be measures of "inhibition" because there are a number of situations in which the two measures change in opposite directions as a function of changes in the same independent variable.

To give one example, consider a small, foveally presented test field that is shown first against a completely dark background and then against a surrounding area (inducing field) whose luminance is approximately the same as that of the test field. Adding the inducing field results in a decrease of the apparent brightness of the test field (Heinemann, 1955), but in an increase in sensitivity (Cobb, 1916; Dittmers, 1920; Steinhardt, 1936). Thus the change in apparent brightness would seem to indicate "inhibition," while the change in sensitivity indicates "facilitation."

Much of what is known of the concurrent variations in apparent brightness and luminance difference sensitivity can be accounted for, in a qualitative way, by the assumption that there is no fixed relation between the level of apparent brightness and sensitivity to luminance differences, but that at any given level of apparent brightness, no matter how produced, luminance difference sensitivity is proportional to the rate of change of brightness with respect to luminance. This, of course, is the heart of Fechner's well-known statement on the subject; though what is usually called "Fechner's Law" is based on some further assumptions that are not made here.

Fechner's Law has been submitted to a variety of tests in the past. A review of many of the experiments in the field of vision may be found in Le Grand (1957). The present report is concerned only with the situation in which the area of the test field, its location on the retina, and the spectral composition of the light are constant, but the test field is subject to the influence of an inducing

[1] This research was supported by a grant from the National Science Foundation.

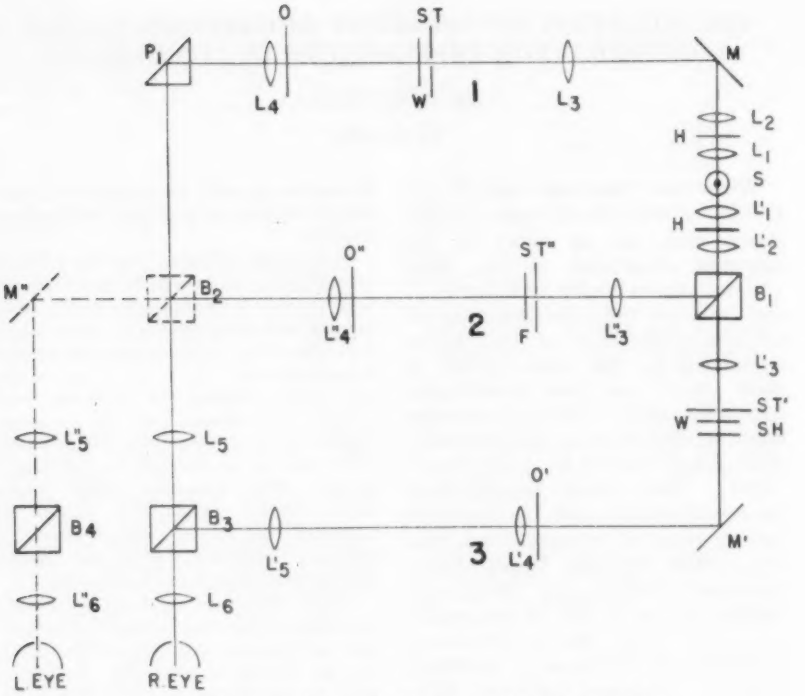


FIG. 1. Diagram of the optical system used in Exp. 1 and 2.

field. For these conditions the applicability of Fechner's Law was investigated by Pitt (1939). It follows from a strict interpretation of Fechner's Law that, whatever the luminance of the surrounding inducing fields, all test fields whose luminance is the same fixed number of just-noticeable-difference (JND) steps above the absolute threshold must have the same apparent brightness. Pitt's experiments showed that this is not true.

Pitt's experiments do not, however, constitute an adequate test of the single assumption concerning the relation between rate of change of apparent brightness and difference sensitivity that was stated above. The reason is that Pitt assumed, in addition, that all stimuli at absolute threshold have the same apparent brightness, "subjective black." This assumption is wrong. Beginning with Hering many investigators have described observations that contradict it. For the particular stimulus situation used in Pitt's experiments, i.e., a stimulus of luminance equal to the absolute threshold presented on a retinal area that is completely surrounded by an inducing field, the present author has shown that the apparent brightness of the threshold stimulus varies with the luminance of the inducing field (Heinemann, 1955).

What follows is a report of two related experiments designed to test the hypothesis that has been discussed. In Exp. 1 measurements were made of variations in the luminance difference threshold in a test field as a function of the luminance

of the test field, for a series of different fixed luminances of the surrounding inducing field. In Exp. 2 a series of functions were determined showing the luminance required to maintain a constant level of apparent brightness in the test field, as a function of the luminance of the surrounding inducing field.

METHOD

Apparatus

Figure 1 is a schematic diagram of the optical system used in both experiments. This optical system could be arranged either for monocular or for binocular viewing. In the monocular version the system does not contain the final portion of the path leading to the left eye, which includes elements M″, L″$_5$, B$_4$, and L″$_6$. S is the source of light, a projection lamp with a horizontal ribbon filament (GE/8A/T10/2P). L$_1$ and L$_2$ are two condensing lenses that collimate the light coming from the rear surface of the ribbon filament. L′$_1$ and L′$_2$ are condensing lenses that collimate the light from the front surface of the ribbon filament. The elements H placed between the condensing lenses are heat absorbing filters.

The light collimated by L′$_1$ and L′$_2$ is partially transmitted and partially reflected by the beam splitting cube B$_1$. There are three beams of light then that are led to *S*'s right eye along three separate paths. To follow one of these, Path 3: the light transmitted by the beam splitting cube B$_1$ next passes through lens L′$_3$. L′$_3$ forms an image of the filament in the plane of ST′. ST′ is a metal stop containing a small circular aperture. W is a circular neutral density wedge. The mounting of this wedge includes a holder for neutral density filters. SH is an electromagnetic shutter. The three elements, ST′, W, and SH are very close together so that effectively all three are in the plane of the filament image. Next the light is reflected by the front surface mirror M′ towards O′ and L′$_4$. O′ is a diaphragm containing an opening whose characteristics determine the size and shape of the field seen by *S*. The lens L′$_4$ forms an image of the filament approximately in the plane of lens L′$_5$. Next the light is reflected by the beam splitting cube B$_3$, passes through the lens L$_6$, and enters *S*'s right eye. An image of the aperture in diaphragm O′ is formed in the focal plane of L$_6$, between B$_3$ and L$_6$, and an image of the filament is formed in the plane of *S*'s pupil. This image of the filament is circular in shape (determined by the opening in stop ST′) and is 1.5 mm. in diameter. Thus *S* has a "Maxwellian view" of the opening in diaphragm O′.

The other optical paths are similar, comparably labeled elements having comparable functions. Differences among the three paths are these: (*a*) Only Path 3 contains a shutter. (*b*) Path 2 contains no wedge but only a holder for neutral density filters, labeled F. (*c*) Path 1 contains P$_1$, a right-angle prism not present in the other paths. (*d*) The light travelling Path 2 passes through one more beam splitting cube (B$_2$) than the light travelling the other paths.

In addition to the aperture used to form part of the basic stimulus pattern to be described below, diaphragms O, O′, and O″ each contain a small opening that is covered by red plastic and illuminated from behind by a miniature bulb. These miniature bulbs are enclosed in small housings attached to the back of these diaphragms. Depending on the requirements of a particular experiment, each of these apertures could be used to form a point of fixation.

For binocular viewing, Paths 1 and 3 remained the same but Path 2 was changed in the following way. The elements L″$_3$, ST″, F, O″, and L″$_4$, were moved to the left by a distance corresponding to the interpupillary distance of *S*. The beam splitting cube B$_2$ was removed from the system, and the first surface mirror M″, lenses L″$_5$ and L″$_6$, and the beam splitting cube B$_4$ were added. The function of the added elements is the same as that of the comparably labeled elements in the other paths. The beam splitting cube B$_4$ has no function other than to equalize the optical elements in Paths 2 and 3.

The *S* sat in a light tight cubicle and his head was held fixed by means of a biting board.

In Exp. 1 the basic stimulus arrangement consisted of three stimuli to be called the "test field," the "inducing field," and the "increment field." The test field was an evenly illuminated disk whose diameter was 30′ of visual angle. Diaphragm O was used to form this test field. The inducing field was an evenly illuminated annulus whose outer diameter was 1° 36′ of visual angle. This annulus surrounded the test field and its inner border was virtually contiguous with the border of the test field. Visually, the test field and inducing field were separated only by a very thin black line. O″, a thin

metal plate, is used to form the annulus. This plate contained a circular aperture that was covered with a thin glass to which had been cemented a thin metal disk.

The increment field was a small disk, 10′ of visual angle, that could be superimposed on the test field. Diaphragm O′ was used to form this increment field. This diaphragm was mounted on one end of a metal rod whose other end was mounted on the voice coil of a loudspeaker. By applying a voltage to the voice coil *E* could move diaphragm O′ a short distance in the horizontal direction, in a plane perpendicular to the optical path. Motion in directions other than that intended was prevented by bearings in which the rod was mounted. This device enabled *E* to present the increment field in one of two positions, with its center either 7′ to the left of the center of the test field or 7′ to the right, by opening and closing a switch. The exposure duration for the increment field was controlled by a modified Gerbrands tachistoscope timer that activated the electromagnetic shutter SH.

In addition to the elements described the stimulus pattern included a small red fixation point placed at the same height as the center of the test field, and, from *S*'s point of view, 52′ to the left of this center.

All the stimuli were presented to *S*'s right eye only. What *S* saw then was a small red fixation point, slightly to the right of which appeared the test field surrounded by the inducing field. A small bright spot, the increment field, could appear briefly on the test field either to the left or to the right of its center.

The stimulus arrangement used in Exp. 2 differed in several ways from that just described. As in Exp. 1, the fixation point, the test field, and the inducing field were presented to *S*'s right eye only. Diaphragm O′ was now used to form the test field, and the slide forming the annulus was placed at O. The increment field was not used in this experiment. Instead, the apparatus was arranged for binocular viewing as described earlier, and a "comparison field" of the same size and shape as the test field was presented to *S*'s left eye only. O″ is the diaphragm used to form the comparison field. The left eye was also presented with a fixation point.

The *S* was required to fuse the two fixation points so that in the fused binocular view he saw a single fixation point to the left of which appeared the comparison field, and to the right of which appeared the test and inducing fields.

As mentioned, the comparison field was a luminous disk of the same size as the test field. It was usually presented against a completely dark background. As long as its luminance is above threshold, such a comparison field appears fairly bright, never dark gray or black. In order to work at levels of apparent brightness lower than those that can be produced in this manner, the comparison field was surrounded by an annulus in several of the experimental sessions. The annulus used was of the same size as the one that surrounded the test field. To produce this disk-annulus arrangement the diaphragm previously used to form the comparison field was replaced by one that contained a circular opening of such size that without further modification it would have formed an evenly illuminated disk whose diameter equalled the outer diameter of the annulus. A small disk of Wratten neutral density gelatin filter, of the same diameter as the test field, was placed between two pieces of cover glass and placed over the opening so that the Wratten-filter disk was centered. This produced the desired stimuli, the luminance of the comparison field depending, of course, on the density of the filter material used.

Subjects

The *S*s were one man and one woman. Each served as *S* in both of the experiments to be described.

Procedure

Experiment 1.—In Exp. 1 the variation in the luminance difference threshold in the test field as a function of the test field luminance was measured at each of six different fixed luminances of the surrounding inducing field. The psychophysical method used was the "up-and-down" method (Dixson & Massey, 1957). The method as used here differed from that described by Dixson and Massey in that "forced choice" was imposed on *S*s. The *S* was always presented with a pair of increment "flashes" delivered 1.5 sec. apart. The duration of each flash was 0.2 sec., and the flashes appeared either to the left or the right of the center of the test field, in a random sequence. After both flashes had been exposed *S* reported on which side of center each had appeared. If *S* reported both positions correctly the luminance of the flash was lowered by 0.1 log unit on the next pair of exposures. If *S* reported one or both of the positions incorrectly, the lumi-

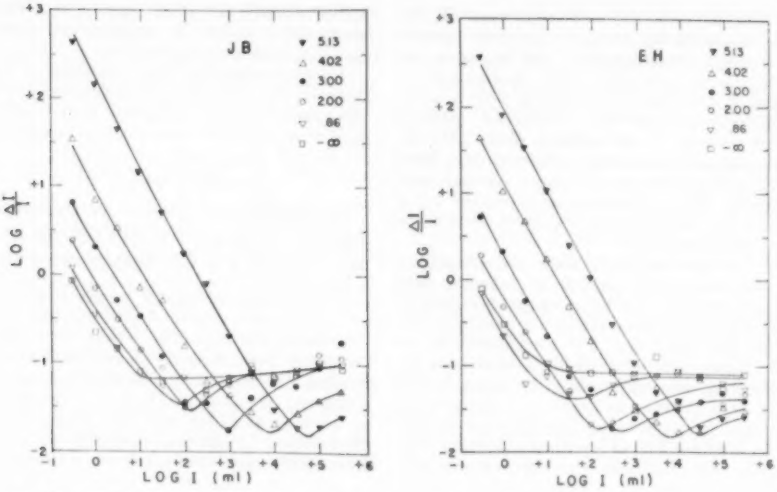


FIG. 2. The log of the Weber fraction as a function of the log test field luminance. The numbers in the upper right corner are the log luminances of the inducing fields.

nance of the flash was raised by 0.1 log unit on the next pair of exposures. The threshold was computed in the manner described by Dixon and Massey (1957). Each threshold is based on a total of 40 judgments.

In a given session, the luminance of the inducing field was held constant at one of the following values: $-\infty$, 0.86, 2.00, 3.00, 4.02, 5.13 log mL., and the difference threshold within the test field was measured at 13 different luminances of the test field, covering the range from -0.54 to 5.46 log mL. in steps of 0.5 log units.

Before each session *S*'s eyes were adapted to darkness for 15 min. Following this *S* was required to stare at the fixation point for about 4 min. During this period, and throughout the experiment the test and inducing fields were exposed continuously.

There were three microswitches within easy reach of *S*. The *S* could expose a pair of the increment flashes by pressing one of these microswitches. The other microswitches were wired to signal lights at *E*'s station and were used by *S* to make his report. The *S* was instructed to make exposures only when *E* signaled by saying "ready." Following the "ready" signal *S* was to make an exposure as soon as he felt that he was fixating well. The *E* gave the "ready" signal approximately every 15 sec. until *S* had made 20 judgments. The *S* then rested for about 4 min. whereupon he resumed the task for a test field of different luminance.

The order of presentation of the 13 test field luminances was always from the lowest to the highest.

This procedure was carried out twice for each of the six inducing field luminances, so that altogether *S* made 40 judgments under each of the stimulus conditions investigated.

Experiment 2.—The basic method used to measure the level of apparent brightness was binocular matching. As mentioned above, the stimulus pattern consisted of two fixation points (one for each eye), the comparison field that was presented to the left eye only, and the test and inducing fields that were presented to the right eye only. By turning a knob that was geared to the proper neutral density wedge *S* could vary the luminance of the test field. He was instructed to vary this luminance until the comparision field and the test field were matched for brightness, taking care to "bracket" the point of equality and to maintain strict fixation while making the adjustment.

In a given session the luminance of the comparison field was held constant at one of eight levels ranging from -0.65 log mL. to $+$ 4.85 log mL. In two of the sessions the comparison field was surrounded by an annulus whose luminance was held constant at $+$ 3.93 log mL.

Before each session *S*'s eyes were adapted to darkness for 15 min. In those sessions in which the comparison field was presented against a completely dark background, *S* began each session by making four matches between the test and comparison fields in the absence of the inducing field. The inducing field was then introduced at a level near threshold and four more matches were made. This procedure was repeated for a series of progressively higher luminances of the inducing field.

In those sessions in which the comparison field was surrounded by an inducing field the comparison field had an apparent brightness so dark that no match was possible unless the test field was also surrounded by an inducing field. The minimum luminance the inducing field surrounding the test field had to have in order to make a match possible was determined in preliminary experiments. An inducing field of luminance slightly higher than this minimum was the first one presented to *S* and, as before, matches were then made for a series of progressively higher luminances of the inducing field.

Each experimental "run" consisting of the determination of one equal brightness curve was repeated once. Altogether then *S* made eight matches under each stimulus condition.

Results

The results of Exp. 1 are shown in Fig. 2.[2] Log $\Delta I/I$ is plotted against log I, where I is the luminance of the

[2] Four tables giving the results of Exp. 1 and 2 have been deposited with the American Documentation Institute. Order Document No. 6609 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

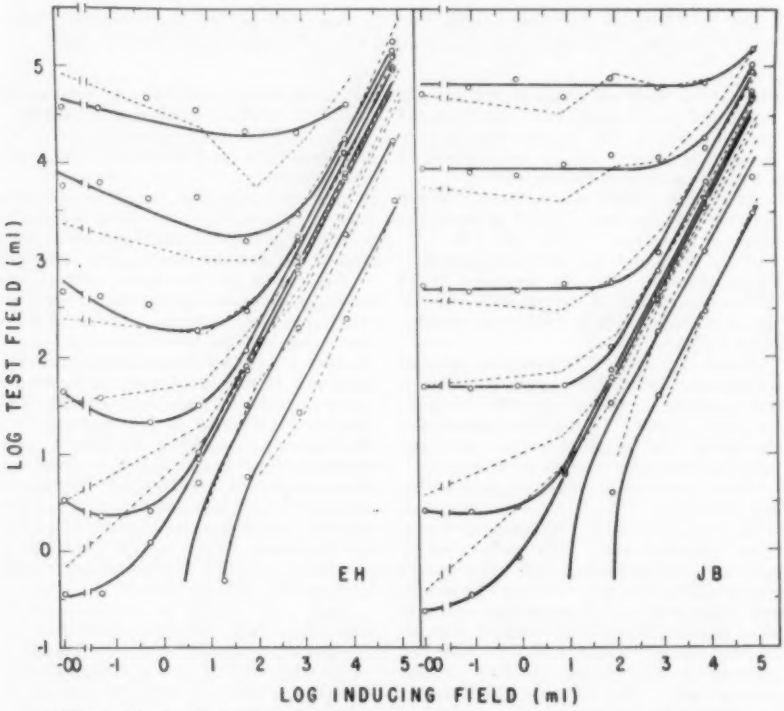


FIG. 3. The log luminance the test field required for constant apparent brightness as a function of the log luminance of the inducing field.

test field and ΔI is the luminance of the increment field at threshold. Each curve is for a different fixed value of the luminance of the inducing field. Curves were fitted to the points visually.

For all the conditions in which the test field was surrounded by an illuminated inducing field the value of log $\Delta I/I$ attains a minimum and then rises again. The position of this minimum depends on the luminances of the inducing field. For the three inducing fields of highest luminance log $\Delta I/I$ has its minimum value when the luminance of the test field is from 0.1 to about 0.5 log units below the luminance of the inducing field. Relative to the abscissa value that indicates the luminance of the inducing field, the position of the minimum progressively shifts to the right as the luminance of the inducing field decreases. For the dimmest inducing field used log $\Delta I/I$ attains its minimal value when the luminance of the test field is almost a log unit higher than the luminance of the inducing field.

Figure 3 shows the results of Exp. 2. The mean log luminance of the test field required to match a comparison field of constant brightness is plotted against the log luminance of the inducing field, for each of eight different apparent brightness levels of the comparison field. (Only the data points, open circles, and the solid curves represent results of Exp. 2. The dashed lines also shown will be discussed below.) The curves were fitted to the points visually. These curves are "constant brightness contours," i.e., the points on these curves represent pairs of test and inducing field luminance values for which the test field has a constant apparent brightness.

For the six curves placed highest in the coordinate system the luminance levels of the comparison fields (corrected for differences in the response of the two eyes) are given by the ordinates corresponding to log inducing field luminance $= -\infty$. For the two lowest curves the luminance of the comparison field is given by the ordinates corresponding to log inducing field luminance $= +3.93$.

In all essential aspects these results are similar to those obtained in an earlier investigation by the author (Heinemann, 1955). When the luminance of the inducing field is considerably below L_0, where L_0 is the luminance the test field requires for a match in the absence of an inducing field, the inducing field either has virtually no effect on the test field (Subject JB) or it may slightly raise the apparent brightness of the test field (Subject EH). As the luminance of the inducing field approaches equality with L_0, it begins to depress the apparent brightness of the test field. For inducing field luminances greater than L_0 this depression, as measured by the increase in the log test field luminance required to maintain a constant level of apparent brightness, is a linear function of the log inducing field luminance.

The two lowest curves represent results obtained with comparison fields that were themselves surrounded by an inducing field of luminance $+3.93$ log mL. The bottom ends of these curves are known to turn so as to become parallel to the y-axis. This was demonstrated in the earlier investigation cited (Heinemann, 1955) in which many measurements were made under the conditions represented by the very bottom portions of these curves.

In order to test the hypothesis discussed in the introductory section the results were

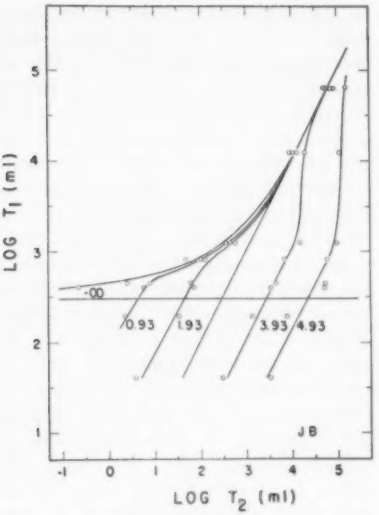


FIG. 4. The relation between the log luminances of two test fields of equal apparent brightness that are surrounded by inducing fields of different luminance. The luminance of the inducing field surrounding T_1 is 2.93 log mL. The parameters on the curves are the luminances, in log mL, of the inducing field surrounding T_2.

treated in the following fashion. First a cumulated JND function was constructed on the basis of each of the curves shown in Fig. 2. To do this, JNDs were added up by a graphical method similar to that described by Luce and Edwards (1958). The number of JND steps between any two test fields of different luminance (but surrounded by the same inducing field) could be obtained from these cumulated JND functions. Next, one of the constant brightness curves shown in Fig. 3, to be called the "base curve," was arbitrarily chosen and a new curve was constructed in the following fashion. To each ordinate of the base curve that corresponds to one of the six inducing field values for which JND data were available was added a quantity representing a constant number of JND steps. Specifically, the quantity added to each ordinate, measured in JND steps, was equal to the mean number of JND steps separating the base curve from a second arbitrarily selected constant brightness curve. This second curve will be called the "target curve." For Subject JB the target curve is the one that intercepts the y-axis at $y = +0.39$, for Subject EH it is the curve that intercepts the y-axis at $y = +0.52$. The six new points thus obtained then lay on a curve whose vertical distance from the base curve, measured in JND units, was the same at all values of the abscissa. Under the terms of the hypothesis that is under consideration this new curve, ideally, should coincide with the target curve.

This procedure was repeated seven times. Each time a different one of the constant brightness functions shown in Fig. 3 was used as the base curve, but the target curve remained the same. This procedure then yielded seven estimates of the shape and position of the target curve. The means of the seven values obtained for each point are plotted on Fig. 3. The six points are connected by straight, dashed, lines to indicate the general trend. The function under discussion is the fourth one from the bottom in each graph.

Finally, the basic procedure described was repeated using these six points, each of which is the mean of seven estimates, as the base curve with each of the seven constant brightness curves not previously so used serving in turn as the target curve. The dashed lines in Fig. 3 represent the family of curves generated in this fashion. It is important to note that the six points predicted for each target curve are connected by straight lines simply to make it easy to see the general trends. The initial segments of these functions almost certainly are curved in the same way as the constant brightness curves. From the manner in which these dashed-line functions were constructed it follows, of course, that they are separated from each other vertically by a constant number of JND steps at all actual values of the abscissa.

Too much significance should not be attributed to precise details concerning the shape and location of the dashed-line functions, for these depend somewhat on how the smooth curves fitted to the data in Fig. 2 are drawn. The inaccuracies of the graphical methods used in the further treatment of the results represent another source of variability. To a reasonable first approximation, in any case, the two sets of curves shown in Fig. 3 appear to represent the same family.

In order to demonstrate some further relationships that are important to the discussion which follows, the results of the brightness matches shown in Fig. 3, for Subject JB, are

plotted in a somewhat different way in Fig. 4. In Fig. 4 the log luminance of a test field, T_1, that is surrounded by an inducing field of +2.93 log mL. is plotted against the log luminance of a second test field, T_2, that has the same apparent brightness as T_1. Each curve then is a locus of points for all of which the two test fields are matched in apparent brightness. The parameters on the curves are the log luminances of the inducing fields that surround T_2. A constant level of apparent brightness is represented by a horizontal line in this plot. The horizontal line that is shown indicates approximately the lowest level of apparent brightness that a test field seen against a completely dark background can have under the conditions of the present experiments. The straight line of slope 1.0 shows the results expected when T_1 and T_2 are surrounded by inducing fields of the same luminance.

Discussion

Figure 4 shows more directly than does Fig. 3 the effect the inducing field has on the rate at which the brightness of a test field changes with changes in its luminance. By definition, the slope of the functions at any point gives the rate of change of the log luminance of T_1 with respect to the log luminance of T_2. But it can be shown that this slope also represents the ratio of the rates of change of brightness of the two test fields (with respect to log luminance). To show this it must be assumed that the brightness of each test field is a function of its luminance, and that the derivatives of these functions exist at all points. One can then write:

$$B_1 = f(\log I_1) \qquad [1]$$

$$B_2 = g(\log I_2) \qquad [2]$$

where B_1 is the brightness of T_1, I_1 is the luminance of T_1, B_2 is the brightness of T_2, and I_2 is the luminance of T_2. Differentiation of these functions yields:

$$\frac{dB_1}{d(\log I_1)} = f'(\log I_1) \qquad [1a]$$

$$\frac{dB_2}{d(\log I_2)} = g'(\log I_2) \qquad [2a]$$

Dividing (2a) by (1a) results in

$$\frac{dB_2}{dB_1}\frac{d(\log I_1)}{d(\log I_2)} = \frac{g'(\log I_2)}{f'(\log I_1)} \qquad [3]$$

If $dB_1 = dB_2$ then $d(\log I_1)/d(\log I_2)$ represents the slope of an equal brightness curve such as those in Fig. 4, and, conversely, if the ratio of the independent differentials $d(\log I_1)$ and $d(\log I_2)$ represents the slope of an equal brightness curve then the dependent differentials dB_1 and dB_2 are equal. For the conditions here under consideration then the dBs disappear from Equation 3. In this form Equation 3 is the statement that was to be derived.

The data presented in Fig. 3 show that, to a reasonable first approximation, the constant brightness functions shown are separated vertically by a constant number of JND steps at all luminance values of the inducing field. In other words, if several test fields are matched for apparent brightness, whether or not they have the same luminance, they remain matched when the luminance of each is changed by the same amount as measured in JND step units.

Now consider again the curves shown in Fig. 4. The statement made in the previous paragraph means that for any two points that lie on the same curve, the vertical distance between these points is equal to the horizontal distance between them, if both distances are measured in JND step units. Starting from this and following Fechner's assumptions in moving from finite differences to differentials, it is easy to demonstrate that the following relationship holds:

$$\frac{dI_1}{dI_2} = C\frac{\Delta I_1}{\Delta I_2} \qquad [4]$$

Where I_1 is the luminance of test field T_1, I_2 is the luminance of test field T_2,

ΔI_1 is the luminance increment threshold in test field T_1, ΔI_2 is the luminance increment threshold in test field T_2, and C is a constant.

In words, for two test fields that are matched in brightness the relative values of the difference thresholds depend on the rate of change of luminance of one of the test fields with respect to the luminance of the other.

There is also a simple relation between the relative rates of growth of brightness in the two test fields and the relative size of the difference thresholds. If I is used as the independent variable rather than log I the same argument used to derive Equation 3 shows that

$$\frac{dI_1}{dI_2} = \frac{g'(I_2)}{f'(I_1)} \qquad [5]$$

From Equations 4 and 5 together it follows that for two test fields that are matched in brightness the ratio of the difference thresholds is inversely proportional to the ratio of the rates of change of brightness with respect to luminance.

To summarize the outcome of both experiments in very general terms: When a foveally presented test field is surrounded by an inducing field that inducing field has three effects. (*a*) The inducing field alters the apparent brightness of the test field. (*b*) The inducing field alters the rate at which the apparent brightness of the test field changes as a function of changes in test field luminance. (*c*) The inducing field alters the luminance difference threshold within the test field. At a given level of apparent brightness, the size of the luminance difference threshold is determined by the rate of change of brightness with respect to luminance.

Returning briefly to a matter that was mentioned in the introductory section; the results under discussion are obviously relevant to the problem of what constitutes a proper index of "inhibition" or "facilitation." If it is assumed, as it often is, that apparent brightness varies directly with the level of activity in the stimulated portion of the nervous system, and if the terms "inhibition" and "facilitation" refer to decreases and increases in this level of activity resulting from the presence of other stimuli, then, clearly, increases and decreases in the luminance difference threshold do not indicate, respectively, "inhibition" and "facilitation." To the contrary, when the sensitivity in an area improves as a result of the stimulation of neighboring areas this is almost invariably the consequence of "inhibition." More precisely, it is a consequence of the manner in which the amount of inhibition exerted by an inhibiting stimulus of fixed luminance varies with the luminance of the "inhibited" area. Analogously, "facilitation" may result in decreased sensitivity.

The statements made apply, of course, only to the type of stimulus situation investigated here. There are many reasons for believing that they do not apply, without modification, to situations in which the light from the several test fields is not of the same spectral composition, or the test fields are of different sizes, or fall upon different regions of the retina, etc. On the other hand, it seems unlikely that the basic principle involved is restricted in its applicability to the one stimulus situation that was investigated. Exploration of a variety of other stimulus conditions is now under way.

Summary

Two experiments were reported. In Exp. 1 six functions were determined showing how the luminance difference threshold within a 30′, foveally presented test field varies with the luminance of the test field. Each function is for a different constant value of the luminance of an annular inducing field that surrounded the test field. The results show that, whenever the test field is surrounded by an illuminated annulus, the value of log $\Delta I/I$ reaches a minimum at a value of log I that depends on the luminance of the inducing field. Beyond this point, log $\Delta I/I$ rises again.

In Exp. 2 eight functions were determined showing, for the same stimulus pattern used in Exp. 1, how the luminance the test field requires in order to maintain a constant level of apparent brightness varies with the

luminance of the surrounding annular inducing field. These functions were determined by the method of binocular matching. Each of the eight functions is for a different level of apparent brightness.

The same *Ss* were used in the two experiments. An analysis based on the results of both experiments shows that (*a*) test fields that are matched for apparent brightness, whether or not they have the same luminance, remain matched when the luminance of each is raised or lowered by the same amount, as measured in JND step units; (*b*) for test fields of the same apparent brightness, the relative values of the difference thresholds depend on the relative rates at which the apparent brightness of these test fields changes with their luminance.

References

Cobb, P. W. The effect on foveal vision of bright surroundings. Part IV. *J. exp. Psychol.*, 1916, **1**, 540–566.

Diamond, A. L. Foveal simultaneous brightness contrast as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1953, **45**, 304–314.

Dittmers, F. Über die Abhängigkeit der Unterschiedsschwelle für Helligkeiten von der antagonistischen Induktion. *Z. Sinnesphysiol.*, 1920, **51**, 214–232.

Dixson, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1957.

Fry, G. A., & Alpern, M. The effect of a peripheral glare source upon the apparent brightness of an object. *J. Opt. Soc. Amer.*, 1953, **43**, 189–206.

Heinemann, E. G. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1955, **50**, 89–96.

Hess, C., & Pretori, H. Messende Untersuchungen über die Gesetzmässigkeit des simultanen Helligkeitkontrastes. *Graefes Arch. Ophthal.*, 1894, **40**, 1–24.

Holladay, L. L. The fundamentals of glare and visibility. *J. Opt. Soc. Amer.*, 1926, **12**, 271–319.

Le Grand, Y. *Light, colour and vision.* New York: Wiley, 1957.

Luce, R. D., & Edwards, W. The derivation of subjective scales from just noticeable differences. *Psychol. Rev.*, 1958, **65**, 222–237.

Pitt, F. H. G. The effect of adaptation and contrast on apparent brightness. *Proc. Phys. Soc.*, 1939, **51**, 817–830.

Schouten, J. F., & Ornstein, L. S. Measurements on direct and indirect adaptation by means of a binocular method. *J. Opt. Soc. Amer.*, 1939, **29**, 168–182.

Steinhardt, J. Intensity discrimination in the human eye: I. The relation of $\Delta I/I$ to intensity. *J. gen. Physiol.*, 1936, **20**, 185–209.

Stiles,. W. S. The effect of glare on the brightness difference threshold. *Proc. Roy. Soc. Lond., Ser. B*, 1929, 104, 322–351.

(Received March 31, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 400–409

TEMPORAL ASPECTS OF CUTANEOUS INTERACTION WITH TWO-POINT ELECTRICAL STIMULATION [1]

ETHEL SCHMID

Veterans Administration Hospital, West Haven, Connecticut

In recent work on vibratory stimulation of the arm, von Békésy (1955, 1957, 1959) found that considerable spread of a mechanical stimulus on the skin does not prevent a sharp localization of the stimulating object. He postulates a neural inhibitory mechanism which counteracts the effects of the physical spread.

Cutaneous inhibitory[2] effects have, in fact, been demonstrated in physiological experiments, although no systematic study of their relation to tactile localization has yet been attempted. In an early study of evoked potentials at somesthetic cortex, Marshall, Woolsey, and Bard (1941) showed that a continuous air blast to one skin locus could mask the response evoked by transient stimulation at another locus. These effects could be obtained from stimulation of widely separated areas, such as the thumb and face. More recently, Mountcastle (1957) discovered that light pressure on the skin surrounding the peripheral excitatory field of a cortical cell inhibited that cell's response to stimulation of its peripheral excitatory field; pressure elsewhere produced no such effects. Amassian (1952, 1953, 1954) and Towe and Amassian (1958) have shown inhibition in various studies of cortical somatosensory Areas I and II and of the somatic "association" area which Amassian discovered.

Considering these physiological experiments and von Békésy's recent work, it is surprising that there is no undisputed demonstration of tactile inhibition in the older psychophysical literature. An early study by Heymans (1901) showed that high intensity stimulation of one skin locus raised the absolute threshold at a neighboring locus or decreased the apparent intensity of a definitely supraliminal stimulus applied at the neighboring locus. But in a series of papers describing their work, von Frey and his colleagues questioned the existence of a "true" tactile inhibitory effect. Working on the forearm, Brueckner (1901), a student of von Frey, reported that he usually obtained summation when two neighboring skin points were stimulated. In a considerable number of cases, however, one of the two stimuli either disappeared or seemed weakened; the latter effect tended to occur particularly when the intensities at the two areas were grossly dissimilar. Brueckner mentioned an apparent spatial displacement of the weaker of the two stimuli toward the stronger. Cook and von Frey (1911) showed a loss of two-point discrimination when one of the two stimuli was made considerably less intense than the other. In cases where two-point discrimination was still possible, *Ss* reported that the weaker stimulus seemed to be drawn toward the stronger. The authors believed that loss of two-point acuity due to spatial displacement of the weaker stimulus toward the stronger, rather than an actual decrease in the apparent intensity of the inhibited stimulus, accounted for any previous reports of inhibition. In order to minimize the effects of apparent spatial displacements on intensity judgments, von Frey and Pauli (1913) performed a series of experiments on the fingers, where the displacement effect is much less pronounced than on the forearm, which was the locus

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Pure Science of Columbia University, 1959.

The author wishes to express very special thanks to Burton S. Rosner for guidance and encouragement throughout the research reported here. Many thanks are due also to Truett Allison and Bettie Nims for their kind cooperation and patience as *Ss*, and to Robert Nims, who administered the anesthetic in the "novocain" experiments, and to Leonard Matin for help in designing the constant current stimulator.

This research was supported in part by Grant M-1530 from the National Institute of Mental Health to Burton S. Rosner.

[2] The word "inhibition" will be used to refer to any decrement in physiological response or increase in psychophysical threshold resulting from the application of a stimulus in another skin area. Nothing about mechanism is implied.

of earlier experiments. They found summation in 16 out of 17 experiments. In the one case where inhibition occurred, the stimulus whose apparent intensity decreased was close to threshold and the intensity of the other stimulus was high. These were the conditions which Cook and von Frey had found to be favorable for the spatial displacement effect.

von Békésy's recent work and the physiological literature demonstrating cutaneous inhibition encourage a re-examination of the problem of psychophysical inhibitory effects during two-point stimulation. No data on the role of a temporal separation between the stimuli are available for the case where a stimulus of fixed intensity (the conditioning stimulus) is applied at one locus and the threshold is determined at a neighboring locus with test stimuli. Accordingly, time between conditioning and test stimuli was the main independent variable of the present experiment. Two additional parameters, intensity of the conditioning stimulus and spatial separation between the conditioning and test stimuli, were also explored. Electrocutaneous stimulation with discrete, brief pulses was used, since it permitted accurate and convenient control of the temporal characteristics and intensity of the stimuli. Absolute thresholds were measured on the third (middle) or fourth digits of the left hand; they were compared with thresholds on those fingers when conditioning shocks were delivered to the second (index) digit. The latter thresholds will be referred to as "conditioned" thresholds.

METHOD

Electrical and electronic equipment.—Two Grass S-4-C stimulators provided the basic pulse-forming and timing circuits. They were used with Grass isolation units to prevent interaction between the conditioning and test stimuli through a common ground. Each isolation unit gated a pentode circuit, also isolated from ground, which delivered a constant current pulse of controllable amplitude and 100 microseconds in duration to *S*. The rise time of the pulse was 1.5 microseconds and the decay time was 2.0 microseconds. Every 20 sec., one of the stimulators was operated by a relay activated by an Industrial Timer, Model CM-2. A synchronizing signal from the first stimulator triggered the second one after a preselected time interval. During the 20-sec. period between stimulus deliveries, *E* set the intensity control at the level required for the next stimulus delivery. The relay click, which occurred 1 sec. before the first stimulus, served as a warning signal for *Ss*.

Calibration of stimulus duration, time between conditioning and test stimuli, and intensity of current was performed by measuring the voltage drop across a known precision resistor in the plate circuit of the pentode unit when the stimulator was triggered. A Tektronix Type 532 oscilloscope and a differential preamplifier (Tektronix Type 53/54D) were used in making these calibrations, which were checked daily before each experimental session. Intensity values stated in this report are reliable to ±1%. Duration and delay values are at worst within ±5% of the stated figures.

Electrodes.—The electrodes were small silver cups, 9 mm. in diameter, filled with bentonite paste[3] and fastened to the skin with Kling Konform bandages (Johnson & Johnson Co.). Preliminary work showed that use of adhesive tape or conventional bandages to fasten the electrodes to the skin tended to produce slight blood vessel pulsations which were easily confused with threshold stimuli; the light Konform bandage minimized, and usually eliminated, this difficulty. The cathode of the conditioning stimulus was centered on the first phalanx of the second digit, with the proximal part of the electrode on the outermost edge of the first interphalangeal joint. The anode was centered on the same finger, 1.5 cm. proximal to the cathode. Similar electrode placements for the threshold test stimulus were made on the third or fourth digits. Pilot work showed that most *Ss* tended to move the two fingers apart while making difficult discriminations. These movements produced variability in the threshold measurements and were prevented in the experiment proper by tying the fingers together with a small piece of Konform bandage. The *Ss* were otherwise free to move their hands and arms as desired.

[3] Anhydrous $CaCl_2$—30.3 g.; aqua dist.—45 cc; bentonite—61 g.; glycerin—6.5 cc.

Subjects.—Preliminary data were collected on 12 *S*s. The final experiment was then conducted on 2 *S*s: B.N. (female), a technician in the laboratory where the experiment was performed, and T.A. (male), a graduate student in psychology at Yale University. Both *S*s received approximately 7 wk. of preliminary training.

Psychophysical method.—Two thresholds were determined at each session: an absolute threshold and then a conditioned threshold, for the particular temporal separation between stimuli and intensity of the conditioning stimulus chosen for that session. In both cases, *S* was instructed to concentrate on the finger where the test stimulus was delivered and to report whether he felt anything in that finger. In Exp. A, with *S* B.N., stimuli to the test finger were presented according to the method of constant stimuli, with a total of 15 stimulus presentations at each of five values of intensity. The stimulus presentations were randomized within three blocks. Each of the three blocks contained 25 stimulus presentations, five at each of the values of intensity used. Steps of 0.03 ma. proved convenient for most measurements; in some cases, 0.06 ma.-steps were used. The results of Exp. A showed that 10 presentations at each intensity were adequate to determine thresholds; accordingly, 10 were used in Exp. B and C. It was then frequently possible to obtain conditioned thresholds for two values of the independent variable in one session. The intertrial time was 20 sec. On the average, experimental sessions lasted 75 min. Control experiments showed that no constant error was introduced by the procedure of taking the absolute threshold first and following it with the conditioned threshold.

Experimental designs.—The design of the experiment appears in Fig. 1, 2, and 3. Time between conditioning and test stimuli (Δt) is shown on the x axis. A positive value signifies that the conditioning stimulus preceded the test stimulus, and a negative value means that the test stimulus preceded the conditioning stimulus. Intensity of the conditioning stimulus is the parameter. For Exp. A and B, B.N. was *S* and the test stimulus was placed on the third and fourth digits, respectively, of her left hand. The *S* in Exp. C was T.A. The test stimulus was placed on the fourth digit of his left hand. The conditioning stimulus was applied to the second digit in all cases. All combinations of intensity (17.8, 14.6, 10.9, 7.4, and 3.6 ma.) and Δt (-30, -20, -10, -5, -1, 0, $+1$, $+5$, $+10$, $+20$, $+30$, and $+40$ msec.) were used in Exp. A only. See Fig. 2 and 3 for the combinations used in Exp. B and C.

Since the possibility existed that current from the conditioning stimulus spread to the locus of application of the test stimulus and so produced a change in threshold at the latter region, the following control experiment was performed. The second digit, where the conditioning stimulus was ordinarily delivered, was completely anesthetized with novocain (2% with suprarenia 1:50,000—Winthrop-Stearns). Thresholds were then measured by the method of limits on the third digit of B.N. and the fourth digit of T.A., with and without accompanying 17.8-ma. shocks (the highest intensity used in the main experiments) to the anesthetized second digit. The two conditions were presented in a counterbalanced order, with *S* unaware of the stimulating conditions. The novocain data from B.N. are therefore a control for Exp. A. Controls for Exp. B and C were provided by T.A.'s novocain data. One set of data was obtained from B.N. in a session beginning about 20 min. after the completion of the novocain injection. As that session progressed, the absolute threshold rose and *S* complained of numbness in her fingers. The determinations were therefore temporarily discontinued. A second set of data was obtained later, approximately 2 hr. after the completion of the novocain injection. The second digit was still completely anesthetized at that time. Only one set of measurements was made on T.A., beginning approximately 20 min. after the completion of the novocain injection.

Results

Graphic plots of relative frequency of feeling against the stimulus intensity showed that the data could be reasonably described by the cumulative normal curve. This was confirmed by a χ^2 test of goodness of fit on a random sample of 10% of the data ($P = .44$ for the sum of χ^2). Accordingly, the obtained data were fitted with cumulative normal curves by the method of least squares. The 50% frequency of feeling point was chosen as the threshold.

Figures 1, 2, and 3 summarize the results of Exp. A, B, and C, respectively. $I_{TC} - I_{TA}$ (the difference in

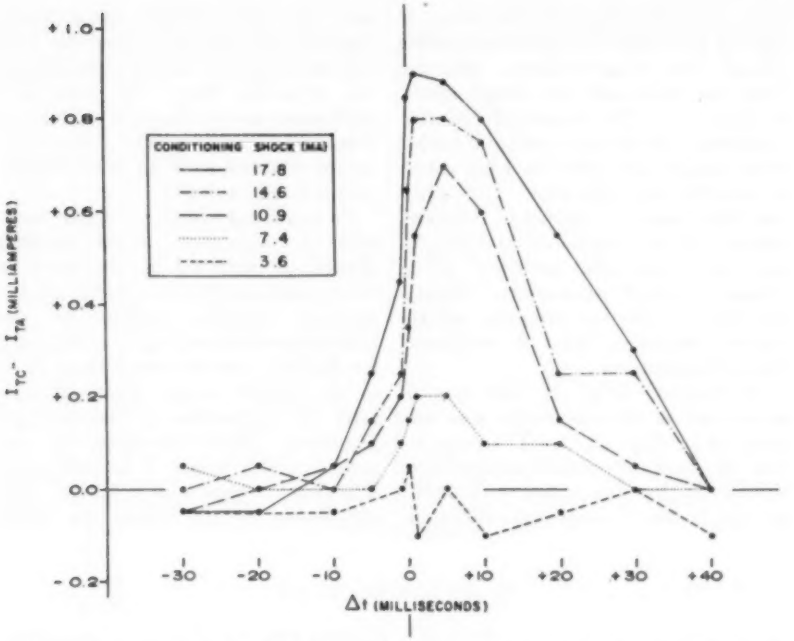


FIG. 1. Results of Exp. A. The difference between the conditioned and absolute thresholds is shown as a function of temporal separation between conditioning and test stimuli. Positive values of Δt signify that the conditioning stimulus preceded the test stimulus. Intensity of the conditioning stimulus is the parameter.

intensity of current between a conditioned threshold and the absolute threshold determined in the same session) is plotted as a function of temporal separation between conditioning and test stimuli. Positive values signify that the conditioned threshold was higher than the absolute threshold. Intensity of the conditioning stimulus is the parameter.

In Exp. A (Fig. 1), the conditioned threshold was higher than the absolute threshold with four out of five intensities of the conditioning stimulus. The amount of inhibition was clearly a function of the temporal separation between conditioning and test stimuli. For a given absolute value of Δt, the effects were greater when the conditioning stimulus preceded the test stimulus than when the test stimulus preceded the conditioning stimulus. By approximately −10 msec. and +40 msec., the conditioned thresholds did not differ significantly from the absolute thresholds. The maximum degree of inhibition occurred at small positive values of Δt (1-5 msec.) and tended to shift toward longer time separations as the intensity of the conditioning stimulus decreased. For the lowest intensity conditioning stimulus used (3.6 ma.) the effects obtained were small and no temporal function was demonstrated. The data suggest possible facilitatory effects with this low intensity conditioning stimulus.

In Exp. B (Fig. 2), the same *S* (B.N.) was used, but the stimuli were placed two fingers apart, whereas they had been only one finger apart in Exp. A. The magnitude of the inhibitory effect was reduced under these conditions. For the high value of conditioning stimulus (17.8 ma.) the data show a temporal function similar to the one found in Exp. A, but the maximum occurred at a longer temporal separation. Again, the data for the low intensity conditioning stimulus show a tendency toward facilitation.

In Exp. C (Fig. 3), the spatial separation of the electrodes was the same as in Exp. B, but T.A. was *S*. The temporal functions obtained were similar to the ones found with B.N. in Exp. B, but the degree of inhibition with the high intensity conditioning stimulus was greater. For the high and intermediate values of conditioning stimulus, there was still some inhibition at the longest values of temporal separation used. No facilitation occurred with the low intensity conditioning stimulus.

Novocain experiments.—Table 1 presents a comparison of the absolute thresholds obtained on the third or fourth digits with and without accompanying 17.8-ma. shocks to the anesthetized second digit. The data for B.N.-I were obtained from B.N. in the session which began 20 min. after the completion of the novocain injection. Data obtained in the session which began 2 hr. after the injection are shown as B.N.-II. The differences in all cases are small

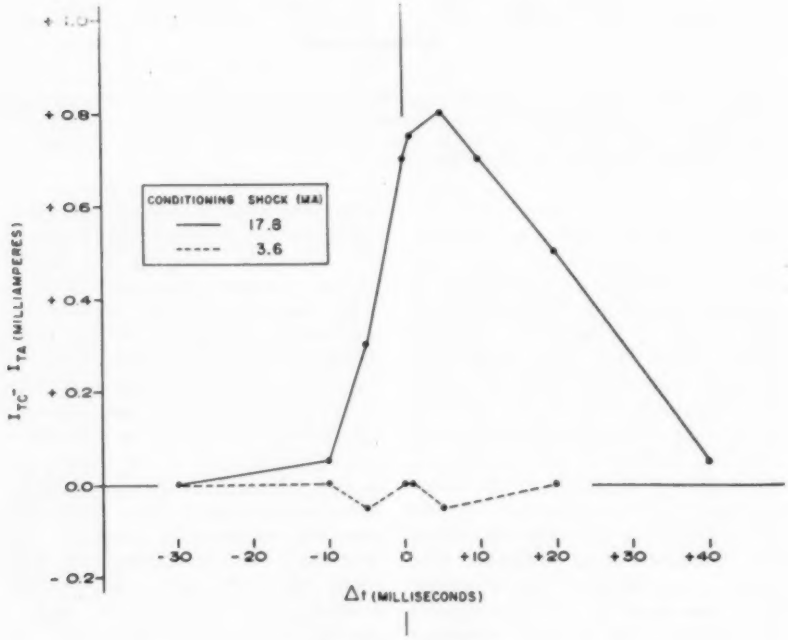


FIG. 2. Results of Exp. B.

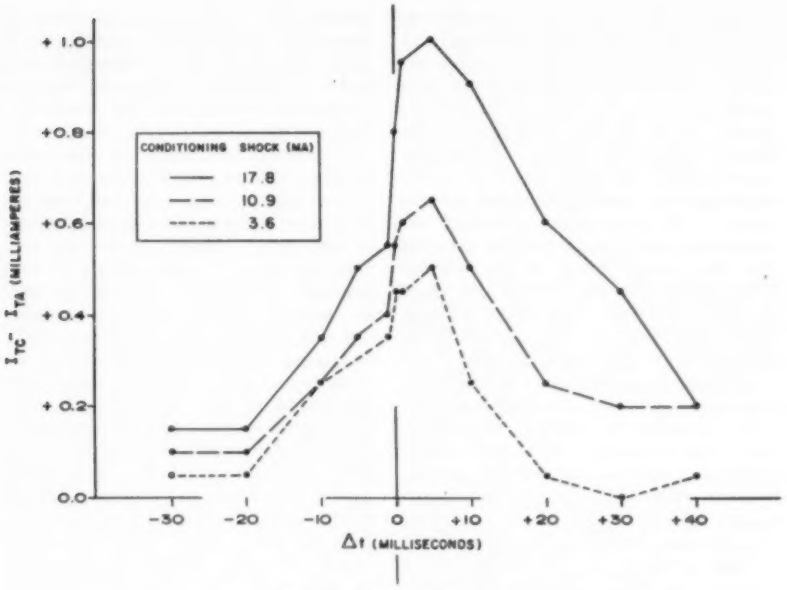


FIG. 3. Results of Exp. C.

(less than 2%), and not statistically significant.

DISCUSSION

For the particular cases shown in Fig. 1, 2, and 3, the conditioning stimulus clearly had a maximum inhibitory effect when it preceded the test stimulus by a small time interval (1–5 msec.).[4] But these results are by no means specific to experiments in which stimulation is applied to the fingers. Pilot data on a variety of skin loci (volar and dorsal surfaces of the forearm; palm; calf; thigh) had shown that a large inhibitory effect occurred between zero and +10 msec.; much less inhibition occurred with the negative Δt values. The fingers were then chosen for the main experiments primarily because preliminary work had shown that it was difficult

[4] The *Ss* were definitely unable to discriminate a temporal separation between the conditioning and test stimuli in the interval between −10 and +10 msec. Under the conditions of this experiment, B.N. could not reliably discriminate a time difference for any of the Δt values used. Subject T.A. was occasionally able to detect a slight temporal separation at the longer values of Δt. In addition to the fact that temporal separations become apparent with Δt values which are longer than the ones used in the present experiment, complex effects, such as apparent movement, occur. The psychophysical discrimination required of *S* under these conditions is not the same as the one required when small temporal separations are used; longer values of Δt were therefore not employed.

TABLE 1

RESULTS OF THE NOVOCAIN CONTROL EXPERIMENTS

Exp.	Absolute Threshold Intensity (ma.)			
	Test Shock and Conditioning Shock		Test Shock Only	
	Mean	σ_M	Mean	σ_M
B.N.—I	1.99	.059	1.97	.037
B.N.—II	1.82	.022	1.84	.009
T.A.	2.06	.016	2.04	.028

to apply high intensity electric stimulation to other skin areas without activating small nerve trunks. After the details of the functions on the fingers were worked out in Exp. A, B, and C, a further check on the generality of the findings was made with a short experiment on the leg of B.N. This study confirmed the results reported for the fingers. In subsequent phases of the discussion, therefore, it will be assumed that the general shape of the function obtained for the fingers will also hold for other skin loci. The point of maximum inhibition, of course, has to be determined for any particular case.

Relation to analagous physiological experiments.—In recent recordings from single neurons in somesthetic cortex, Towe and Amassian (1958) have shown inhibitory effects with stimuli similar to those employed in the present experiment. Shocks delivered to one digit of a monkey inhibited responses of the cortical neurons to stimulation of a neighboring digit. Despite this similarity in stimulating conditions, most of the data they present have a time course radically different from the one found in the present experiment. In one study, for example, they measured probability of neural firing as a function of time by which the conditioning stimulus to the second digit preceded the test shock to the third digit. The results showed that typical cells were not affected at zero temporal separation; a pronounced inhibition occurred between approximately 5 and 90 msec.; this inhibition was followed by a period of facilitation. Amassian (1953) showed similar long-lasting temporal effects in earlier experiments with shocks to nerve trunks.

The discrepancy between these physiological data and the present results does not necessarily imply, however, that there is no relation between the behavioral events and the firing of cortical neurons. First, it is possible that the species difference is an important variable in this case. Secondly, studies of cortical evoked potentials have shown that some anesthetics produce a lengthening of the recovery cycle when two successive stimuli are applied at the same peripheral locus (Marshall et al. 1941). A similar temporal distortion could be operating in Towe and Amassian's work on inhibitory effects, where two different cutaneous loci were stimulated. Their own data lend some support to this notion, since they find that their temporal functions are affected by depth of anesthesia. Thirdly, there are selective factors in present electrical recording techniques which make it impossible to sample cortical neurons randomly (Mountcastle, Davies, & Berman, 1957). It is possible that there are cortical cells which show a time course similar to the one found in the behavioral data, but that their morphology operates against the probability of recording from them. With this point in mind, it is interesting to note that Towe and Amassian mention a few cases (three cells) in which inhibition occurred when conditioning and test shocks were applied simultaneously or even when the test shock preceded the conditioning shock. Amassian (1952) also found evidence of inhibition under these conditions in some recordings of evoked potentials. Unfortunately, however, no detailed functions were published for these cases.

Relation to problem of tactile localization and stimulus spread.—Tactile localization is puzzling from two points of view. Ever since the notion of a small skin area, innervated by a single cutaneous fiber, has been proved incorrect, workers have been puzzled by the precision of localization despite the broad and overlapping receptive fields of single cutaneous fibers. Early thinking on the problem emphasized the importance of patterning in the firing of a group of neurons (Hoagland, 1932; Tower, 1943). More recently, several writers have stressed the role that cutaneous inhibition may play in tactile localization and acuity (Amassian, 1952; Mountcastle, 1957).

The second problem was raised by von Békésy, whose experiments showed traveling waves on human skin which spread several centimeters but were not

recognized by the subject. Von Békésy also suggested that an inhibitory mechanism improves localization. For purposes of the present discussion, it seems worthwhile to formulate this second problem somewhat more explicitly than he has done.

It is apparent that even if a traveling wave may be detected physically over a long distance, the damping could be such that the stimulus is subliminal after a considerably shorter distance. It is, of course, only this shorter distance which logically poses a problem in understanding tactile localization.

Apart from any complex central or peripheral inhibitory mechanism, simple refractoriness in the skin neurons themselves could improve localization. Consider a hypothetical, simplified section of skin (Fig. 4), which contains four neurons, each spreading over a considerable distance and overlapping extensively with other neurons. A transient stimulus is applied at Point X and fires Neurons a and b. Provided that the spread of stimulus is sufficiently rapid, Neuron b will still be refractory when the traveling wave reaches Point Y. Assuming that the psychophysical threshold at a particular skin locus is a function of the *effective* neuron density at that locus, the probability of a response from Point Y is reduced. A similar situation will obtain as the stimulus spreads from Y to Z. The available data on size of receptive fields of single cutaneous fibers and on the rate of travel of mechanical waves over the skin are compatible with these notions.[5] Thus, when we take into account the physical spread of the tactile stimulus, it is apparent that skin with a simple, nonoverlapping afferent supply would actually be poorly equipped for coding information on stimulus location.

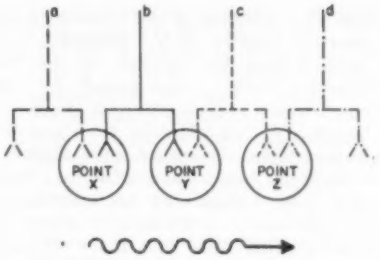


FIG. 4. The diagram represents a small, simplified section of skin, innervated by Neurons a, b, c, and d. A transient stimulus applied at Point X spreads in all directions. For simplicity, only travel horizontally and to the right is considered in the diagram. The effects at Points X, Y, and Z are discussed in the text.

There is, however, definite evidence that refractoriness in peripheral neurons is not the only mechanism involved in cutaneous "inhibition." Much of von Békésy's data, including his demonstration of bilateral inhibitory effects (1955), obviously calls for a different explanation. Furthermore, the comparatively long time course of the inhibitory effect studied in the present experiment points to some mechanism in addition to refractoriness in peripheral neurons. Additional evidence is available from Amassian's work (1952). He reported only a small inhibitory effect in recording from the medial lemniscus; for the same stimulating conditions, he found a much larger effect in thalamic afferents.

Relation to other psychophysical experiments.—In the introduction, early psychophysical experiments were described which showed mutual facilitation between two cutaneous loci except when there were gross intensity discrepancies between the stimuli applied to them. In this latter case, inhibition occurred, which von Frey interpreted as a loss of acuity due to the spatial displacement of the less intense stimulus toward the more intense one.

Evidence of a spatial displacement

[5] Diameters of receptive fields ranging from less than 1 mm. to more than 10 cm. have been reported (Maruhashi, Mizuguchi, & Tasaki, 1952). Rate of travel of a mechanical wave over the skin depends on the underlying tissue. The available data show rates from less than 10 m/sec to over 1000 m/sec (Keidel, 1952).

was also obtained in the present experiment. Subject B.N. frequently remarked that the location of the test stimulus when accompanied by the conditioning shock differed from its location during absolute threshold determinations. It seemed to be displaced toward the conditioning stimulus. Subject T.A. volunteered no information on this matter. When questioned, he reported that the occurrence of such a displacement was by no means clear; if any occurred, the test shock moved toward the conditioning shock. The present observations thus are compatible with von Frey's report of a spatial displacement effect. But clearly, a change in threshold also occurs, and there was no good evidence that this change in threshold was only a by-product of the spatial displacement effect. Using small temporal separations between two equally intense stimuli, von Békésy (1957) also showed that both a spatial displacement and a decrease in apparent intensity were reported. But he found summation between two transient stimuli delivered to the skin simultaneously. The stimuli chosen in the present experiment, of course, emphasized the inhibitory case. Nonetheless, a slight facilitation may have occurred with B.N. when the low intensity conditioning stimulus was used (i.e., in the case where the intensity discrepancy between the two stimuli was less pronounced).

If spatial summation occurred between two skin points under all conditions of stimulation, as von Frey concluded, the problem of stimulus spread and localization would obviously be aggravated. But the available data really only show summation between two simultaneously applied transient stimuli when their intensities are not greatly disproportionate. Inhibition occurs when intensities are disproportionate or when there are small temporal separations between stimuli. When a mechanical wave, initiated by a transient stimulus applied at a single locus, travels over the skin, the effective stimulating conditions are such that we could expect suppression of the sensation in the regions to which the wave spreads: there is a temporal separation between stimulation at the center of an area and at the border, and the stimulus at the border is less intense than the stimulus at the center. Simultaneous, equally intense stimulation of two separate points, however, produces a mutual facilitation. This mutual facilitation, combined with inhibition in the areas between the two stimulated points, could obviously be a mechanism for enhancing two-point discrimination.

Summary

The effect of a conditioning electric shock at one skin locus on the threshold for electrocutaneous stimulation at a neighboring locus was studied as a function of the temporal separation between the two stimuli (Δt). Results generally showed an inhibitory effect which was maximal when the conditioning shock preceded the test shock by a short temporal interval (1–5 msec. for the particular stimuli used). For a fixed value of Δt, the amount of inhibition was a monotonic function of the intensity of the conditioning shock. The inhibition was greater when the two stimuli were placed closer together spatially. In one *S*, a slight facilitation occurred when a low intensity conditioning stimulus was used.

Results of the study are discussed in relation to relevant physiological experiments, the psychophysical literature on the problem and recent biophysical experiments on traveling waves on the skin. Their significance for the problems of tactile localization and two-point acuity is considered.

References

Amassian, V. Interaction in the somatovisceral projection system. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1952, **30**, 371–402.

Amassian, V. Evoked single cortical unit activity in the somatic sensory areas. *EEG clin. Neurophysiol.*, 1953, **5**, 415–438.

Amassian, V. Studies on organization of a somesthetic association area, including a single unit analysis. *J. Neurophysiol.*, 1954, **17**, 39–58.

Brueckner, A. Die Raumschwelle bei Simultanreizung. *Zeit. Psychol. Physiol. Sinnesorgane*, 1901, **26**, 33–60.

Cook, H., & von Frey, M. Der Einfluss der Reizstärke auf den Wert der simultanen

Raumschwelle der Haut. *Zeit. Biol.*, 1911, **56**, 537–573.

HEYMANS, G. Untersuchungen ueber psychische Hemmung. *Psychol. Physiol. Sinnesorg.*, 1901, **26**, 305–382.

HOAGLAND, H. A basis for cutaneous localization. *J. gen. Psychol.*, 1932, **6**, 468–470.

KEIDEL, W. Messung der Hautwellengeschwindigkeiten bei Vibrationsreizen am Menschen. *Pflug. Arch. ges. Physiol.*, 1952, **255**, 213–227.

MARSHALL, W., WOOLSEY, C., & BARD, P. Observations on cortical somatic sensory mechanisms of cat and monkey. *J. Neurophysiol.*, 1941, **4**, 1–24.

MARUHASHI, J., MIZUGUCHI, K., & TASAKI, I. Action currents in single afferent nerve fibers elicited by stimulation of the skin of the toad and the cat. *J. Physiol.*, 1952, **117**, 129–151.

MOUNTCASTLE, V. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.*, 1957, **20**, 408–434.

MOUNTCASTLE, V., DAVIES, P., & BERMAN, A. Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. *J. Neurophysiol.*, 1957, **20**, 374–407.

TOWE, A., & AMASSIAN, V. Patterns of activity in single cortical units following stimulation of the digits in monkeys. *J. Neurophysiol.*, 1958, **21**, 292–311.

TOWER, S. Pain: Definition and properties of the unit for sensory reception. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1943, **23**, 16–43.

VON BÉKÉSY, G. Human skin perception of traveling waves similar to those on the cochlea. *J. Acoust. Soc. Amer.*, 1955, **27**, 830–841.

VON BÉKÉSY, G. Sensations on the skin similar to directional hearing, beats and harmonics of the ear. *J. Acoust. Soc. Amer.*, 1957, **29**, 489–501.

VON BÉKÉSY, G. Similarities between hearing and skin sensations. *Psychol. Rev.*, 1959, **66**, 1–22.

VON FREY, M., & PAULI, R. Die Stärke und Deutlichkeit einer Druckempfindung unter der Wirkung eines begleitenden Reizes. *Zeit. Biol.*, 1913, **59**, 497–515.

(Received March 31, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 410-416

PREDICTION OF RECURRENT SEQUENCES AS RELATED TO LEVEL OF IRRELEVANT CUES[1]

ABRAM M. BARCH

Michigan State University

Event-prediction experiments (sometimes called verbal conditioning or probability-learning experiments) have generally varied only the first-order probabilities of the events to be predicted. Some investigators have studied predictive success when higher-order probabilities have been varied (Bennett, Fitts, & Noble, 1954; Hake & Hyman, 1953; Nicks, 1959). Very few event-prediction studies have used a recurrent fixed sequence (Bruner, Wallach, & Galanter, 1959; Galanter & Smith, 1958; Goodnow & Pettigrew, 1956; Namikas & Brogden, 1960) although interest in such problems is not new, as shown by the long standing work on double alternation (e.g., Gellerman, 1931). The learning of sequences is an integral aspect of the learning of serial verbal lists and mazes of various kinds. But such situations commonly have a distinct beginning and a distinct end for the series or some point in the sequence consistently associated with a reinforcement while recurrent sequences do not. The purpose of the present experiment is to study predictive success on two different types of recurrent fixed sequences as related to the level of irrelevant cues.

Several investigators have pointed out the difficulty of obtaining a quantitative measure of pattern characteristics. Thompson, Voss, and Brogden (1957) attempt merely to note the effect of doublets—and split doublets—at different positions in a verbal maze. The results of Galanter and Smith (1958) show little promise in the use of series length or of total information value (bits). Nicks (1959) proposes the description of sequence structure in terms of number and length of event runs.

Since quantitative description of sequences appears difficult in the present state of our knowledge, use was made of qualitative distinctions believed to be of importance. Therefore, two types of sequences were developed which differed with respect to run characteristics, expected sensitivity to negative recency effects (Jarvik, 1951), and symmetry, but which had constant first-order probabilities, length, and number of different combinations of successive events. These sequences are described in detail in the Method section. A negative recency tendency is here regarded as a tendency to avoid choosing event (stimulus) classes recently seen.

The term "irrelevant cue" (or dimension) has various interpretations in the experimental literature. The usage here is the same as that of Archer (1954). A relevant cue or dimension is one that can be used by *S* to make a correct prediction (or classification); an irrelevant cue is one that varies from trial to trial but cannot be used to make a correct prediction. A cue which does not vary is neither relevant nor irrelevant.

No data are available on the effect of varying level of irrelevant information on prediction of a recurrent event

[1] The work reported here was supported by a research grant (G-3338) from the National Science Foundation. The assistance of Thomas Trabasso and Ronald Hoppe is gratefully acknowledged.

sequence. After the present study was completed, Bruner, Wallach, and Galanter (1959) presented a theoretical schema of identification of recurrent regularity in which they proposed that accurate recognition of environmental regularity would be impeded by presence of inappropriate information (stimulus interference). Such stimulus interference could result from the failure of the sequence to maintain complete regularity or from variation of irrelevant aspects of the stimuli. They reported evidence supporting their hypothesis but only tested the effect of introducing randomness into the recurrent sequence.

METHOD

Subjects.—The *Ss* were 196 college men and 71 college women from the introductory psychology course, Winter of 1957, who received research credit for serving. Of these, 25 *Ss* were eliminated (19 men and 6 women) for failure to follow instructions, apparatus failure, or *E*'s error. One to four *Ss* served in a single experimental session; all *Ss* at a session received the same experimental treatment with treatments being assigned to sessions in a predetermined order. To equate the size of the groups and to make the sex proportions as similar as possible among the groups, eight additional *Ss* were eliminated at random. The number of *Ss* remaining in each group was 39 with the number of women ranging from 7 to 15 per group.

Procedure.—The *Ss* were told that their task was to guess which of four forms (triangle, circle, cross, and star) would next be projected from film-strips by a PerceptoScope onto a 2 × 2 ft. wall-mounted screen which was 13 ft. from the projector. The *Ss* had 5 sec. in which to make each guess. Guesses were made by depressing one of four keys on a labeled keyboard. The *Ss* were seated 8 ft. from the projection screen and their view of each other was blocked by screens placed between them.

Each *S* was given four practice guesses (trials) in which each form appeared once and 192 learning trials. The *E* recorded each correct guess. A 1-min. rest was given between the practice and learning trials and after each block of 48 learning trials.

Tasks.—The two types of fixed sequence are designated as A and B. The A (or symmetrical) sequence consisted of: triangle, star, circle, cross, cross, circle, star, triangle, repeated in that order six times in each block of 48 trials. The B (or restricted) sequence consisted of triangle, star, circle, triangle, cross, circle, star, cross, repeated in that order six times in each block of 48 trials. Note that six positions of these eight-position sequences are identical. They differ only at Positions 4 and 8.

Sequence A has runs of two crosses and two triangles; Sequence B has no form immediately succeeding itself. Each has the same number of higher order groups (eight) for each group length (e.g., triangle, star or circle, star, cross, triangle). Sequence A is symmetrical, with the second half of the sequence the mirror image of the first. Sequence B has an obvious lower level of symmetry.

In Sequence B, once it is underway, all odd-numbered positions (P-1, P-3, P-5, P-7) contain events (forms) which have not been seen in the previous three trials while the even-numbered positions contain events not seen in the previous two trials. If a negative recency tendency dominated, each of the odd-numbered positions should be predicted better than each of the even-numbered positions of this sequence.

Once Sequence A is underway, P-4 and P-8 contain events not seen in six previous trials; P-3 and P-7 contain events not seen in four previous trials; P-2 and P-6 contain events not seen in two trials, while P-1 and P-5 contain events seen on the immediately preceding trial. If negative recency can be regarded as a graded response set, with the tendency to avoid an event class decreasing with increased remoteness, then the predictive success of the pair of positions should be in the same order as described.

Three levels of irrelevant cues were used. The relevant dimension of form always had an information value of 2 bits. Groups A-0 and B-0 saw only one red-colored form on each frame of the film-strip, and thus had 0 bits of irrelevant information. The forms in the film-strips for Groups A-2 and B-2 varied in a restrictive random fashion over four colors (blue, green, red, or yellow) for 2 bits of irrelevant information. Groups A-4 and B-4 saw sequences which varied in number of forms in each frame (from one to four) as well as in color, and thus contained 4 bits of irrelevant information. The restriction on randomness for the irrelevant dimension was that all colors or numbers appear

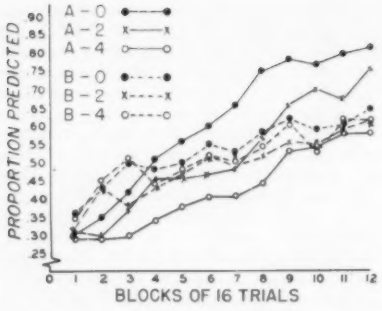


FIG. 1. Mean proportion correct over blocks of 16 trials.

equally often in 64 trials. The stimuli used were the entire deck of the Wisconsin Card Sorting Test (Grant & Berg, 1948).[2]

Two different forms were never presented in the same frame. Each series of 48 frames was preceded and followed by a neutral gray frame.

RESULTS

Figure 1 presents the mean proportion correct for blocks of 16 trials each for the six experimental groups. It is obvious that the variation in level of irrelevant cues has a considerable effect on the learning of groups with Sequence A. All groups tend to start slightly above chance, probably owing to prior preferences with respect to associations between the forms. Note the surge during the first 32 trials in Groups B-0 and B-4.

Analyses of variance (Lindquist Type III, 1953) were applied to the initial prediction scores (first 48 trials), final prediction scores (last 48 trials), and the difference, for each *S*, between initial and final scores.

A single overall analysis including all stages of practice was not performed because both the group means and the group variances increased considerably from the first to the last 48 trials. There was no relation, however, between group means and variances within either stage of practice or within the gain measures. The analysis of difference scores permits an unambiguous test of slope, or rate of performance change.

The analysis for the initial stage of practice is presented in Table 1. The effect of type of sequence (ST) was highly significant ($F = 34.16$, $df = 1/228$) with Sequence B yielding a significantly higher proportion of correct predictions than Sequence A. There was also a significant interaction between the type of sequence and level of irrelevant cues (IC) ($F = 4.48$, $df = 2/228$). But the group variances were significantly heterogeneous ($P < .01$). Therefore, the significant ST × IC interaction should be interpreted with caution.

The proportion correct for each of the groups is given in Table 2. Groups B-4 and B-0 scored significantly higher than all the Sequence A groups and Group B-2 was higher than Group A-4 ($P = .05$). The sequential Q technique tabled by Snedecor (Snedecor, 1956, p. 251)

TABLE 1

ANALYSIS OF VARIANCE OF MEAN PROPORTION CORRECT IN FIRST 48 TRIALS

Source	*df*	*MS*	*F*
Between *Ss*	233		
Sequence type (ST)	1	3.621	34.16**
Irrelevant cues (IC)	2	.239	2.25
ST × IC	2	.475	4.48*
Error (between)	228	.106	
Within *Ss*	1638		
Positions (P)	7	.825	18.33**
P × ST	7	1.142	25.37**
P × IC	14	.088	1.96*
P × ST × IC	14	.061	1.36
Error (within)	1596	.045	

* $P = .05$.
** $P = .01$.

[2] Copies of the Wisconsin Card Sorting Test were provided through the courtesy of David A. Grant.

TABLE 2

MEAN PROPORTION CORRECT FOR SEQUENCES A AND B FOR THE FIRST AND LAST 48 TRIALS

Irrelevant Cues	Sequence A		Sequence B	
	First	Last	First	Last
0	.35	.80	.43	.61
2	.33	.71	.37	.59
4	.29	.57	.44	.59

was used to make all individual comparisons.

The various positions in the series yielded different prediction levels and this factor interacted significantly with type of sequence. The proportion correct for each position for Sequences A and B is given in Table 3.

The odd-numbered positions in Sequence B were all significantly higher ($P = .05$) than the even-numbered positions of Sequence B and all positions of Sequence A except P-4, supporting a negative recency interpretation.

In general, Sequence A predictions also followed negative recency expectations: P-4 and P-8 were predicted the best; P-7 and P-3 were next in predictive success; P-6 follows the rule but P-2 does not, being too low; P-5 and P-1 are the lowest except for P-2. P-5, however, is already showing some signs of overcoming the negative recency tendency.

Despite the fact that the two sequences are alike for six of the eight positions, the predictive success for the same position varies widely from Sequence A to Sequence B during the first 48 trials with each position significantly different from its corresponding position in the other series.

The P × IC interaction was borderline and not especially revealing. P-2 was significantly lower than the other positions despite the variation in level of irrelevant cues. On the other hand, certain positions in the 0-bits groups, such as P-5 and P-7, were favored.

For the final stage of practice, both sequence type ($F = 12.05$, $P < .01$) irrelevant cues ($F = 6.70$, $P < .05$) and the ST×IC interaction ($F = 4.59$, $P < .05$) were significant. The mean proportion correct for the groups for the final stage of practice is given in Table 2. Group A-0 was significantly higher than all other groups. Group A-2 was significantly higher than all others except Group A-0.

The various positions were significantly different ($F = 27.23$, $P < .01$) and interacted with sequence type again ($F = 26.59$, $P < .01$). No other source was significant. The mean proportion for each position for the two sequences for the last 48 trials is given in Table 3.

Once again the odd-numbered Sequence B positions are significantly higher than the even-numbered B positions. On the other hand, learning by *Ss* of the doublets in Sequence A has reduced the possibility of ranking positions on the basis of negative recency. If P-1 and P-5 are ignored, the negative recency

TABLE 3

MEAN PROPORTION CORRECT FOR SEQUENCES A AND B FOR EACH POSITION FOR THE FIRST AND LAST 48 TRIALS

Position	Sequence A			Sequence B		
	Event	First	Last	Event	First	Last
1	Triangle	.25	.69	Triangle	.53	.69
2	Star	.19	.51	Star	.29	.55
3	Circle	.33	.69	Circle	.49	.68
4	Cross	.45	.81	Triangle	.30	.45
5	Cross	.33	.78	Cross	.50	.70
6	Circle	.30	.59	Circle	.39	.55
7	Star	.34	.72	Star	.48	.67
8	Triangle	.41	.75	Cross	.32	.48

hypothesis still ranks positions correctly in terms of predictive success.

Comparing the same positions across the two sequences, only Positions 4, 5, and 8 are significantly different and the difference for Position 5 is marginal. Even after considerable practice the second cross of the doublet was anticipated only slightly better than it would have been without a doublet arrangement while the second of the triangle doublet showed no superiority at all. The effect of doublet learning for the triangle would be expected to be less than for the cross since the 1-min. rest, when it came, separated the two triangles.

All groups showed significant gains from initial to final stage of practice. Other results from the gain scores were similar to the analysis for the final 48 trials, with the exception of a nonsignificant interaction between ST × IC. The Sequence A groups gained significantly more than the Sequence B groups; groups with 0 and 2 bits of irrelevant information gained more than the groups with 4 bits. The 0-bits groups just failed to score significantly higher than the 2-bits groups.

Discussion

At first glance, the extent to which the results support a thorough application of a negative recency principle is surprising. From the findings of binary choice (such as Jarvik, 1951, or Nicks, 1959) one might expect a positive recency effect for short runs and a negative recency effect for runs of three or more. Goodnow (1955) infers from her protocols that her *S*s expected runs of two or three. On the other hand, Thompson, Voss, and Brogden (1957), using a four-choice verbal maze, found significantly more errors on the second item of a doublet or on the third item of a split doublet than on the first. Also, Bennett, Fitts, and Noble (1954) in a five-choice event prediction study found definite tendencies to avoid immediate repetitions of a choice.

The available data do not allow us to pinpoint the difference between these situations. However, Goodnow (1955) has proposed that *S*s base their choices on run structure while Nicks (1959) finds that *S*s respond to alterations in run structure even when first-order probabilities remain unchanged.

We propose that *S*s enter situations which they regard as "guessing" or chance situations with expectations related to run length. They expect longer runs with few alternatives and shorter runs with many alternatives. Thus, avoidance of a repetition of choices recently made should be stronger in series with a larger number of alternatives.

After exposure to the series, they modify this expectation in the direction of the run-structure experienced for the various alternatives, with some generalization of run-structure expectation from alternative to alternative. Minor modifications may be introduced in the fine structure by the correctness of an immediately preceding prediction. If this interpretation is correct, negative recency effects of the extent obtained here do not require a fixed sequence.

Goodnow and Postman (1955) and Goodnow (1955), in studies intended to test a proposed distinction between chance and skill situations, report the absence, as hypothesized, of negative recency effects in an event-prediction study presented as a problem-solving situation. However, their problem-solving task contained irrelevant cue variation from trial to trial, thus raising the possibility of a confounding of skill and irrelevant information. The strong presence of negative recency effects in the present study, which varied irrelevant cues, lends indirect support to a guessing-skill distinction.

The initial inferiority of Sequence A groups also agrees with the finding of Ernst, Hoffeld, Sidenstein, and Brogden (1960) that the occurrence of a doublet in a four-choice verbal maze pattern

has an initial inhibitory effect upon acquisition followed by a tendency to facilitate learning later in practice. It should be noted that Voss and Ziegler (1960) found this effect stronger with two doublets than with one.

Bennett, Fitts, and Noble (1954) studied triad, or trigram, learning (where the prediction must be made on the basis of the two preceding events) using probabilistic sequences and found no evidence for learning. Predictive success on our Sequence B increased significantly from the first block of trials to the last even though this improvement required that *S* learn triads or longer series.

The Sequence B and conditions of administration were designed to exclude a number of factors that Bennett, Fitts, and Noble had proposed to differentiate their task from fixed sequences or probabilistic sequences of other types where learning was found. Here, first-order probabilities were equal; the sequences were not grouped spatially, temporally, or in terms of previously learned meaningful units. Further, the time between individual events was greater than in their study and a much shorter total amount of practice was given.

The higher-order probabilities were not equal. But, the most significant difference seems to be in the number of different combinations of events. Fixed sequences will always have a considerably fewer number of possible combinations of events than probabilistic ones.

The Bruner, Wallach, and Galanter (1959) hypothesis of retardation of identification of regularity was supported by the results of the Sequence A groups but not by those of Sequence B. There is a possibility that extended practice might have resulted in a significant effect. Nevertheless, the influence of irrelevancy seems to be dependent on the difficulty level of the sequence itself. One tenable hypothesis is that level of cue irrelevancy is potent only within some middle range of task difficulty.

Summary

The acquisition of predictive success on two different recurrent sequences of events was studied in relation to the level of irrelevant cues. Each eight-unit sequence was composed of four geometric forms equally represented. Sequence A had two runs of two each; Sequence B had runs of only one. Sequence B permitted partial solution by use of a negative recency tendency; Sequence A tended to penalize such a tendency. Color and number of forms provided 0, 2, and 4 bits of irrelevant information. Six groups received four blocks of 48 prediction trials each. Prediction of Sequence B was superior initially but inferior later and showed no effect of variation in irrelevant information. Prediction of Sequence A was superior later, especially with lower levels of irrelevant information.

References

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48,** 313–317.

Bennett, W. F., Fitts, P. M., & Noble, M. The learning of sequential dependencies. *J. exp. Psychol.*, 1954, **48,** 303–312.

Bruner, J. S., Wallach, M. A., & Galanter, E. H. The identification of recurrent regularity. *Amer. J. Psychol.*, 1959, **72,** 200–209.

Ernst, R. L., Hoffeld, D. R., Sidenstein, S., & Brogden, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze patterns. *J. exp. Psychol.*, 1960, **59,** 94–103.

Galanter, E. H., & Smith, W. A. S. Some experiments on a simple thought problem. *Amer. J. Psychol.*, 1958, **71,** 359–366.

Gellerman, L. W. The double alternation problem: II. The behavior of children and human adults in a double alternation problem. *J. genet. Psychol.*, 1931, **39,** 359–392.

Goodnow, J. J. Response-sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68,** 624–630.

Goodnow, J. J., & Pettigrew, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51,** 385–392.

Goodnow, J. J., & Postman, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49,** 16–22.

Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and a case of shifting to new responses in a Weigl-type card-sorting problem. *J. exp. Psychol.*, 1948, **38,** 404–411.

Hake, H. W., & Hyman R. Perception of the statistical structure of a random series

of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

NAMIKAS, G., & BROGDEN, W. J. Effect of sequential patterns upon serial-position errors and acquisition of a verbal maze. *J. exp. Psychol.*, 1960, **60**, 50–56.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

SNEDECOR, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

THOMPSON, R. F., Voss, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

Voss, J. F., & ZIEGLER, J. A. Effect of number of doublets upon verbal maze learning. *J. exp. Psychol.*, 1960, **59**, 182–184.

(Received April 1, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 417–421

EFFECTS OF VARIOUS ASYMPTOTIC RESTRICTIONS ON HUMAN TRIAL-AND-ERROR LEARNING[1]

RIDGELY W. CHAMBERS[2] AND CLYDE E. NOBLE

Montana State University

The present research program in human learning is attempting to extend Hull's S-R reinforcement theory (Hull, 1952) from the special case of two-choice simple trial-and-error learning to the general case of multiple-choice compound trial-and-error learning. Noble (1957a) has proposed the following rational equation as a working model to describe the acquisition curves:

$$R_p = a(i)r^N \qquad [1]$$

where R_p = probability of correct choice (based on the pooled responses of a group of *Ss*); a = asymptote or limit of R_p (reflecting the proportion of reinforceable responses); i = initial probability at the outset of learning (given by the reciprocal of the number of independent and mutually-exclusive available responses); r = rate parameter (theoretically related to stimulus conditions, task factors, work distribution, reward contingencies, and individual differences); and where N = number of practice trials.

Previous experiments have investigated the generality of the equation under variations of response availability (Noble, 1955, 1957c), task length (Noble, 1957b), locus of reward (Noble & Noble, 1958), delay of reward (Noble & Alcock, 1958), specificity of instructions (Noble, Alcock, & Farese, 1958), type of choice procedure (Noble, Alcock & Noble, 1958), individual differences (Noble, Noble, & Alcock, 1958), work distribution (Noble & Taylor, 1959), and uniformity of the stimulus sequence (Farese & Noble, 1960). The percentage of variance in R_p accounted for by Equation 1 in these studies has ranged in 24 independent samples from 91.8% to 99.9% with the higher values predominating (mean = 98.7%).

The present experiment is a more stringent test of the applicability of Equation 1 in which the limit of acquisition (a) was varied for different groups, all other variables being kept constant. Various asymptotic restrictions were introduced by manipulating the proportion of reinforceable responses available to *Ss* in a standard 10-link problem. The empirical question, in other words, was to determine the influence of different "degrees of solubility" on the basic form of human trial-and-error acquisition curves.

Most of the earlier investigations found a high degree of predictability for Equation 1 when i was varied and a held constant. This experiment was designed to assess the predictability of the formula under exactly the opposite conditions. Our expectation was that Equation 1 would continue to forecast human behavior with approximately the same accuracy as before, and that r would be some function of a. Due to the nature of the equation, it was also deduced that original proficiency of different experimental groups would vary directly with these asymptotic restrictions, i.e., when $N = 0$, $r^N = 1.00$

[1] This paper is based upon a master's thesis performed by RWC under the guidance of CEN. We are grateful to Hyrel M. Chambers and Quentin A. Leland for aid in data analysis; and to Robert B. Ammons and Howard E. Reinhardt for helpful comments. The research was supported in part by the National Science Foundation under Research Grant G4895. An abridged report was made to the Midwestern Psychological Association, St. Louis, Missouri, April 1960.

[2] Now at the University of Arizona.

hence $R_p = a(i)$ before training has begun.

Method

Apparatus.—The Selective Mathometer, described in detail elsewhere (Noble, 1957a; Noble & Farese, 1955) is an automatic 19-key pushbutton device designed for the study of complex learning in human *Ss*. Pacing stimuli are presented serially by an electronically-controlled slide projector, to which *S* responds by selecting keys on a semicircular panel. The slides consisted of an invariant sequence of 10 electrical circuit symbols, and the correct solution for all *Ss* was to press the odd-numbered keys in the order: 3, 13, 17, 9, 19, 5, 1, 11, 7, 15. Scores were recorded by polygraph and counters. Time constants for stimulus duration (2 sec.), intraserial interval (1.75 sec.), interserial interval (7.50 sec.), reward delay (0 sec.), and reward duration (.75 sec.) were the same as in most previous experiments. All data were collected in an air-conditioned laboratory which isolated *E* and *S* in adjoining soundproofed chambers. The rooms were connected by a one-way vision mirror and an intercommunication system.

Subjects.—The *Ss* were 150 students enrolled at Montana State University during the spring and summer quarters of 1959. The sample was composed of 74 men and 76 women, whose ages ranged from 17 to 59 yr. (mean = 25.18 yr.). None had previous experience with the Mathometer. All received standard instructions (Noble, Alcock, & Farese, 1958, p. 255) which emphasized accuracy, quick responding, and the dependency of the green light on correct choices. The nature of the apparatus and the instructions were such that *Ss* received no advance indication of differential treatment in the various conditions.

Procedure.—The *Ss* were divided into five groups of 30 each and assigned to the different conditions by a simple forward-reverse counter-balancing plan. The treatments differed in the fact that the number of reinforceable keys varied from 4 to 10. The reference or control group (Group 10) practiced under standard conditions where 10 keys were available and all were rewardable, thus making it possible for *Ss* in this group to find the appropriate key for each stimulus. Group 10 thus worked on a *soluble* problem, where $a = 1.00$. Group 9 was presented with the same stimuli and had the same keys available as Group 10, but one of the keys was nonreinforceable; i.e., one key always failed to produce reward or feedback, making it possible for *Ss* in this group to find an appropriate key for only 9 of the 10 stimuli. Group 9 therefore worked on an *insoluble* problem, where $a = .90$. For Group 8 the situation was the same as above except that two keys were nonreinforceable; hence $a = .80$. Group 6 had four keys nonreinforceable, giving $a = .60$. Finally, Group 4 practiced with six nonreinforceable keys, so $a = .40$.

The four groups whose asymptotic values were less than 1.00 were further subdivided into three equal subgroups ($n = 10$) which were differentiated from each other by the pattern of nonreinforceable keys. In all cases the keys were eliminated by random selection as shown in Table 1. The noncorrection procedure (Noble, Alcock, & Noble, 1958) was used throughout, with *S* being allowed only one choice for each slide. All *Ss* were instructed to use the forefinger of the preferred hand only, to return the hand to the fixation plate after each response, and not to press any buttons during the interstimulus interval. Each *S* practiced for 50 trials or mastery, whichever came first, the latter being defined as five successive trials without error, a criterion which had been established in a previous study (Noble, 1957b).

TABLE 1

KEYS ELIMINATED

Groups	Subgroup 1	Subgroup 2	Subgroup 3
9	9	17	3
8	9, 5	9, 15	1, 7
6	3, 13, 17, 9	3, 13, 9, 15	3, 9, 11, 15
4	3, 19, 1, 11, 7, 15	9, 19, 5, 1, 11, 15	3, 13, 9, 5, 11, 15

Results and Discussion

The acquisition curves are presented in Fig. 1, where the probabilities of correct (reinforceable) responses (R_p) are plotted as a function of trials (N) for the five values of the asymptote parameter, a. It should be emphasized that each curve contains relative frequency measurements of only those acts which are both correct ($R+$) and rewardable (G), and that the denominator (number of possible responses) for each trial score is 10. As predicted from theoretical con-

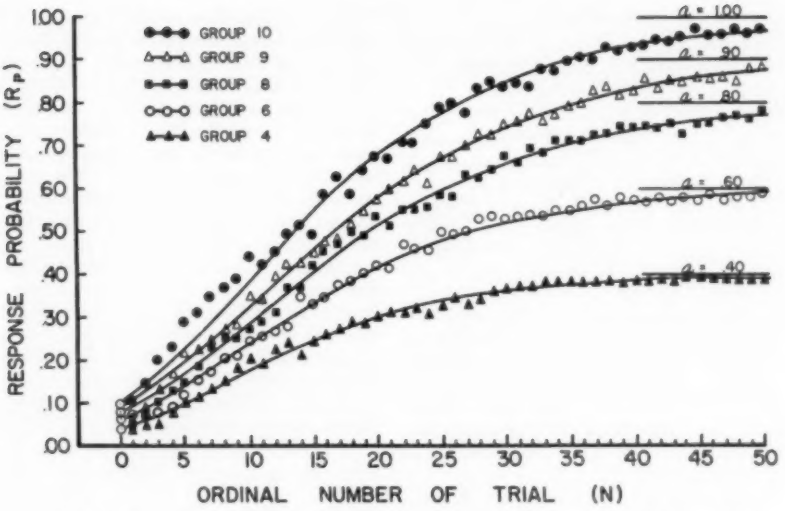


FIG. 1. Compound trial-and-error learning (R_p) as a function of practice (N) with different asymptotes (a) as the parameter. The origins are rational.

siderations (Hull, 1952; Noble, 1957a), all five curves are asymmetrical and sigmoidal. Moreover, they are in excellent agreement with Equation 1. The best-fitting theoretical equations, obtained by minimizing the sums of the squares of the residuals to determine the rate parameter, r, are as follows:

$$\text{Group 10:}\quad R_p = 1.00(.10)^{.916N} \qquad [2]$$
$$\text{Group 9:}\quad R_p = .90(.10)^{.922N} \qquad [3]$$
$$\text{Group 8:}\quad R_p = .80(.10)^{.922N} \qquad [4]$$
$$\text{Group 6:}\quad R_p = .60(.10)^{.913N} \qquad [5]$$
$$\text{Group 4:}\quad R_p = .40(.10)^{.902N} \qquad [6]$$

The computational procedure for finding r has been fully described elsewhere (Noble, 1957b). A preliminary estimate of the inflection point of the curve (N_i) was made by dropping a perpendicular from a point on the ordinate equal to the ratio a/e (where $e = 2.718$) to the abscissa. This approximate N_i value was then substituted in the formula,

$$\log_e r = \frac{-\log_e(-\log_e i)}{N_i} \qquad [7]$$

to obtain an estimate of r. The exact value of r is secured when the quantity $\sum(R_p - \bar{R}_p)^2$ reaches a minimum. In all previous studies, the ordinate at which acceleration of the R_p curve passed to zero was a constant; namely, $.368 = 1/e$, the reciprocal of the base of the natural system of logarithms. A salient feature of the present experiment is the demonstration that this interesting property of Equation 1 is merely the *special* case of a soluble problem where the asymptotic value is 1.00. In the *general* case of problems ranging in degree of solubility (probability of mastery = a) from zero to unity, the inflection point should occur at a/e. Empirical confirmation of this prediction can be seen in Fig. 1.

Coefficients of determination for Equations 2 through 6 were computed directly from the variances of the residuals and the R_p values, as recommended earlier (Noble, 1957b). Predictability is over 99% in every case. As a test of our hypothesis that r would be a function of a, the rate parameter of best fit for Group 10 is compared with each of the others in Table 2. Judging by the

TABLE 2

RATE PARAMETER OF GROUP 10 COMPARED WITH RATE PARAMETERS OF BEST FIT FOR THE OTHER GROUPS

Group	r	Goodness-of-Fit Index (%)	$r = .916$
10	.916	99.06	
9	.922	99.57	98.59
8	.922	99.41	98.58
6	.913	99.31	98.46
4	.902	99.15	95.17

relatively greater amounts of variance in R_p scores accounted for by Equations 3, 4, 5, and 6 when their least-squares r values are employed rather than an r of .916, it is clear that the rates of change are systematically, although not monotonically, affected by manipulation of the asymptotes.

Zero accelerations occur at the following trials (N_i) for the five fitted acquisition curves: Group 10: 9.5; Group 9: 10.27; Group 8: 10.27; Group 6: 9.1; Group 4: 7.58. Examination of the empirical points in Fig. 1 indicates that the R_p values for Group 10 deviated markedly from its theoretical curve. A z test of proportions taken at the point of maximal deviation (Trial 5) indicates that a deviation ($R_p - \hat{R}_p$) this large or larger would occur 38.98% of the time by chance. The Trial 5 discrepancy, therefore, is not significant.

As a further test of the differential-rate hypothesis, a 10×5 mixed-factorial analysis of variance was performed on the $R+$ scores, arranged in 10 blocks of five trials (N) for the five experimental groups (a). The main effects of N ($F = 92.90$, $df = 9/1305$) and a ($F = 80.36$, $df = 4/145$) are both significant. As statistical evidence of nonparallel trends, the significant $N \times a$ interaction ($F = 1.78$, $df = 36/1305$) also supports our expectation that r would vary as a function of a.

It is of some interest to evaluate the effects of degree of solubility on the *difficulty* (Noble, 1957b) of acquiring comparable amounts of "knowledge" (number of R+'s) about the task or problem. To this end a trials-to-criterion analysis was made of the five groups in which the criterion was defined as that trial on which the same four correct-and-rewarded responses were made for the fifth trial in succession. Figure 2 shows the median number of trials required to reach this criterion ($R_{\dot{N}}$) as a function of the percentage of reinforceable responses ($100a$). The relationship is an inverse proportionality whose best-fitting, least-squares equation is $R_{\dot{N}} = -.176(100a) + 35.924$. Calculation of the coefficient of determination indicates that 96% of the variance in $R_{\dot{N}}$ is attributable to $100a$. From this we conclude that difficulty is a negative linear function of increasing solubility over the present range.

Further examination of Fig. 1 indicates that the initial proficiency (R_p on Trial 1) of the different groups varies directly with the asymptotic restrictions. By reference to Equations 2 through 6 it is evident that the theoretical origin of each group's curve (R_p on Trial 0) is given by the product of a and i. This important characteristic of Equation 1 may be rationalized by applying the multiplication theorem of the probability calculus, as follows. When 10 keys are available, the probability of getting an R+ by chance alone (not considering reward contingencies) is $.10 (= i)$. When the number of reinforceable responses is restricted, however, we are dealing with the combined probability of independent events, each less than unity. For this reason, we must

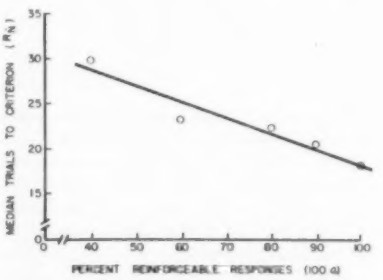


FIG. 2. Difficulty in learning four choices ($R_{\dot{N}}$) as a function of degree of solubility ($100a$) of the 10-link problem. The fitted line is empirical.

calculate the joint probability that a particular key will be selected ($R+$) *and* that this response will be followed by reinforcement (G). In abstract terms: $p(R+\bigcap G) = p(R+)\cdot p(G)$. In the case of Group 4, for example, $p(R+)=i=.10$ and $p(G) = a = .40$; hence R_p on Trial $0 = a(i) = .04$. The joint probabilities of the other groups are computed in similar fashion. Thus is the rationale of Equation 1 directly coordinated with the relative frequency theory of probability.

In conclusion, it appears from these results that Equation 1 has again proved itself to be an excellent predictor of human acquisition data in complex learning situations. The versatility of the equation is indicated not only by its ability to describe the empirical results with an error of less than 1%, but also by the extent to which it lends itself to systematic formulation in probability terms.

Summary

This experiment investigated the effects of asymptotic restrictions on the forms of acquisition curves in a complex trial-and-error learning problem in man. Five groups of 30 *Ss* each practiced a series of 10 pushbutton responses on the Selective Mathometer for 50 trials. The number of reinforceable keys was manipulated in the five groups to produce 40%, 60%, 80%, 90%, and 100% limits on the attainable levels of mastery.

The response probability (R_p) curves were asymmetrical and sigmoidal when plotted as a function of amount of practice (N). Each group's curve originated at a hypothetical R_p value given by the combined probability of making a correct choice (i) and of receiving reinforcement (a). The rational equation $R_p = a(i)r^N$ accounted for over 99% of the variance in the dependent variable scores (R_p) in all five instances. The rate parameters (r) were systematically, although not monotonically, affected by variation of the asymptotes. It was shown that in general the inflection points of the theoretical curves occurred at ordinate values of a/e, where $e = 2.718$.

An analysis of the influence of degree of solubility on difficulty agreed with the effects of asymptotic restrictions on initial proficiency. By measuring the median trials required to master four correct responses in all groups, it was found that difficulty was an inverse linear function of solubility.

References

Farese, F. J., & Noble, C. E. Trial-and-error vs. mixed-selective learning in man. *Percept. mot. Skills*, 1960, **10**, 115–122.

Hull, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

Noble, C. E. Compound trial-and-error learning as a function of response availability (N_R). *J. exp. Psychol.*, 1955, **49**, 93–96.

Noble, C. E. Human trial-and-error learning. *Psychol. Rep.*, 1957, **3**, 377–398. (a)

Noble, C. E. The length-difficulty relationship in compound trial-and-error learning. *J. exp. Psychol.*, 1957, **54**, 246–252. (b)

Noble, C. E. Supplementary report: Response availability in human trial-and-error learning. *J. exp. Psychol.*, 1957, **53**, 359–360. (c)

Noble, C. E., & Alcock, W. T. Human delayed-reward learning with different lengths of task. *J. exp. Psychol.*, 1958, **56**, 407–412.

Noble, C. E., Alcock, W. T., & Farese, F. J. Habit reversal under differential instructions in compound trial-and-error learning. *J. Psychol.*, 1958, **46**, 253–264.

Noble, C. E., Alcock, W. T., & Noble, J. L. The interaction of type of choice procedure with amount of practice in trial-and-error learning under two reward conditions. *J. Psychol.*, 1958, **46**, 295–301.

Noble, C. E., & Farese, F. J. An apparatus for research in human selective learning. *J. Psychol.*, 1955, **39**, 475–484.

Noble, C. E., & Noble, J. L. Human trial-and-error learning under joint variation of locus of reward and type of pacing. *J. exp. Psychol.*, 1958, **56**, 103–109.

Noble, C. E., Noble, J. L., & Alcock, W. T. Prediction of individual differences in human trial-and-error learning. *Percept. mot. Skills*, 1958, **8**, 151–172.

Noble, C. E., & Taylor, A. Influence of work distribution upon complex learning by the noncorrection and modified-correction methods. *J. exp. Psychol.*, 1959, **58**, 352–356.

(Received April 1, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 422–429

INFORMATION CONTENT IN RECOGNITION AND RECALL

R. DAVIS, N. S. SUTHERLAND,[1] AND B. R. JUDD

Oxford University

Wherever recognition and recall have been compared, it has been found that more correct responses are made in recognition than in recall (Postman, Jenkins, & Postman, 1948, and further references cited there). It seems possible that this difference is due to there normally being a greater number of alternatives to choose from in recall than in recognition, and that there may be no fundamental difference between the psychological mechanisms at work behind recognition and recall. If this is correct, the same amount of information should be transmitted in both recognition and recall. The purpose of the present experiment was to test this hypothesis.

Judd and Sutherland (1959) have developed a method of measuring the information content of non-sequential messages, i.e., messages which consist of a series of symbols in which no symbol occurs more than once and no information is transmitted by the order of the symbols. The method allows us to estimate the average information transmitted per *S* when the quantities k, g, q, r are known where k is the total number of symbols in the ensemble from which selection can be made, g is the number of symbols sent, q is the number received, and r is the number correct (i.e., the number which appear both in the set sent and the set received). If recognition and recall are to be compared in terms of information transmitted, it is necessary to select material where the number of alternatives from which selection is made (k) is finite and known; in recognition this number is fixed by the number of alternatives given in recognition lists but there must also be a finite number of alternatives in the case of recall. In previous experiments on recall there has usually been no clear limit to the possible number of alternatives from which selection is made, and therefore this condition has not been satisfied. The material used in this experiment was designed to provide a finite, and manageable number of alternatives in recall in that *S*s had full knowledge of the complete set from which symbols were drawn (90 two-digit numbers or 90 two-letter syllables). The number of symbols sent (g) was kept constant at 15. Since *S*s were allowed to select as many symbols as they chose in recognition and recall, q varied from *S* to *S*. One of the advantages of using an information theory measure in this situation is that it makes possible a meaningful comparison between performances in which the number selected (q) varies.

[1] One of the authors (NSS) wishes to acknowledge financial assistance from the Nuffield Foundation and the American Office of Naval Research. His share of this work was undertaken as part of a project on "stimulus analysing mechanisms" jointly financed by those bodies. We are also grateful to Davidson of the Dorset House School of Occupational Therapy and to Durran of the Oxford Eye Hospital for their co-operation in the experiment and to the *S*s they provided.

METHOD

Subjects.—There were 24 women *S*s between the ages of 18 and 25 yr.; 18 were training to be occupational therapists and 6 were studying orthoptics.

Procedure.—The stimuli used were two-digit numbers and two-letter syllables. There are 90 two-digit numbers, and a list of 90 two-letter syllables was made up by starting each syllable with a consonant omitting q, x, and z and ending each syllable with a vowel. Thus 18 consonants and 5 vowels were used, giving 90 syllables in all. From each list of 90 syllables four lists of 15 stimuli were selected by random sampling without replacement. Thus there were four lists of 15 numbers and four lists of 15 syllables, and no number or syllable appeared in more than one list. Within each list individual numbers and syllables were presented in a random order: the order was the same for each *S*.

Lists were presented in typescript on a memory drum which exposed each symbol for 1.5 sec. The experiment lasted for four days and on each day each *S* had two lists presented, the first a list of letter pairs, the second a list of number pairs. On each day all *Ss* were presented with the same lists. The letter lists were originally regarded merely as providing practice for *Ss* on each condition, but the results for letters were internally consistent and were in most ways similar to the results for numbers, so both sets of data are reported.

There were four conditions: (*a*) Recall: *Ss* were asked to recall as many syllables or numbers as they could; (*b*) Recognition 30: *Ss* were presented with a list of 30 numbers or syllables and had to mark all numbers or syllables which they thought they could recognize as having appeared in the stimulus list; (*c*) Recognition 60; (*d*) Recognition 90. Each list presented for checking in the recognition conditions contained all 15 syllables or numbers which had been presented in the preceding stimulus list. The number recognition lists were arranged in order of ascending magnitude, and the letter recognition lists were arranged in alphabetical order. Recognition lists were typed in the same typescript as the stimulus lists. There are 24 possible orders in which the four conditions can be given successively: a different order was used with each *S*, and each possible order was therefore used once. On each day each *S* had the same condition for both letters and numbers. Thus on Day 1 all *Ss* had the same two lists of 15 numbers and 15 syllables presented to them, but six *Ss* were recognizing out of 30, six out of 60, six out of 90, and six were asked to recall the material; similarly on Day 2, all *Ss* had the same two lists (but different lists from those used on Day 1), and each *S* performed under a different condition from Day 1 once again with six *Ss* under each condition, and so on for the remaining two days of the experiment.

The daily procedure was as follows: first *S* was given instructions relevant to the first list to be presented (letters), she was then shown the list, and recognition or recall of that list was obtained. Instructions concerning the second list (numbers) were then given, the second stimulus list was presented and recognition or recall of that list obtained. The instructions given were as follows: on each trial *Ss* were told whether they would see syllables or numbers. They were also told that the syllables would each be made up of a consonant followed by a vowel and the consonants q, x, z would not appear or that only numbers between 10 and 99 would be shown. On Day 1, it was explained to them approximately how fast the memory drum would move. In the recall condition they were told that they must repeat as many of the syllables or numbers as they could, starting 5 sec. after the disappearance of the last stimulus in response to a signal from *E*. In the recognition conditions they were told that 5 sec. after the disappearance of the last stimulus they would be presented with a list of syllables or numbers. They were told that all 15 of the syllables or numbers they had seen would appear in the list, and that they were to check as many of the syllables or numbers as they could recognize. For both recall and recognition conditions *Ss* were told to "guess whenever you think you may get a number (or syllable) right but are not quite certain, but don't guess wildly." They were also told that the order of recall or the order in which items were checked in recognition was not important: that they could recall or check in any order they chose. Finally they were instructed to make responses as fast as possible in recall since it had been found in a pilot experiment that *Ss* tended to take longer over the last few syllables in recall than in recognition.

Results

The raw scores are set out in Table 1. It is not possible to compare directly the raw scores, e.g., with syllables *Ss* made 181 correct responses out of a total of 303 in recognition out of 90, and it is not possible to say whether this is better or worse than their score of 149 correct out of 207 responses in recall. It is worth

TABLE 1

TOTAL CORRECT RESPONSES, TOTAL ERRORS, AND AVERAGE INFORMATION SCORES (I)

Condition	Day 1			Day 2			Day 3			Day 4			Total	
	Correct	Error	I	Correct	Error	I	Correct	Error	I	Correct	Error	I	Correct	Error
Letters														
Recog. 30	58	18	4.0	61	14	8.1	68	15	8.0	63	20	5.0	250	67
Recog. 60	50	28	8.5	52	23	11.7	41	21	6.6	58	22	12.6	201	94
Recog. 90	48	29	13.6	51	26	16.0	36	37	6.8	46	30	11.2	181	142
Recall	30	21	6.7	39	11	13.6	35	16	9.6	45	10	15.9	149	58
Numbers														
Recog. 30	56	16	4.7	55	12	5.8	58	18	4.7	56	19	4.8	225	65
Recog. 60	44	17	10.5	44	32	6.5	41	20	7.9	50	20	11.0	179	89
Recog. 90	42	28	9.9	31	29	4.8	37	32	7.7	30	37	3.9	140	126
Recall	43	21	11.5	38	16	10.6	39	20	12.3	41	22	10.7	161	79

noting that significantly more responses were given in recognition out of 90 than in recall. Wilcoxon's (1949) test for paired replicates was applied to *S*s' total responses in recognition out of 90 and recall and the difference was found to be significant at better than the .02 level for syllables and better than the .01 level for numbers.

In order to obtain a more meaningful comparison between conditions, the scores of individual *S*s were converted to information scores using Equation 24 from Judd and Sutherland (1959). The mathematical procedure is there set out in detail. Briefly it is assumed that where S makes a given number q of responses, over a series of trials the probabilities of S getting r correct for all possible values of r will be distributed in such a way as to minimize the information conveyed (i.e., will be distributed randomly) given that the average number S will get right is s. If s is known, a minimization procedure can be used to determine information transmitted for different values of k, g, and q. The number of correct responses (r) for each S was used to obtain the best estimate of s under each condition for each S; since the value of k, g, and q is known for each S it is then possible to convert her score into a score in terms of information transmitted. It has been shown that owing to lack of linearity in the I,s curve this procedure overestimates the total information transmitted. However, the amount overestimated tends to be constant, and was corrected for in the calculation by deducting .72 bits. For Recall and Recognition 90, $k = 90$; for Recognition 60, $k = 60$; and for Recognition 30, $k = 30$. The number g was always 15 since it is the number of symbols sent; and the number q for each S was the total number of responses made.

In addition to raw scores, Table 1 shows the average amount of information transmitted (I) per S per day for each condition. Examination of these scores suggests that there is no very marked practice effect from day to day. Practice effects were examined by performing two t tests on the information scores for each of the eight conditions: one test was made to determine whether performance on Day 1 differed from performance over Days 2–4, and the other to determine whether performance on Day 1 differed from performance on Day 4. Only one t test yielded a result significant at better than the .05 level (difference between scores for recall of letters on Days 1 and 4); only two others were significant at better than

the .1 level (differences between scores on recall of letters on Day 1 and Days 2–4, and on recognition of numbers out of 90 on Day 1 and Days 2–4). With the possible exception of the recall of letters, we seem justified in concluding that practice effects had little influence on the information scores, and we shall therefore consider below information scores computed over the whole four days. This does not of course mean that proactive effects do not partially determine the results. Thus it is possible that the effect of "learning to learn" was roughly counterbalanced by proactive inhibition from the similar materials learned on previous days. It will also be noted that scores for the syllables both tend to fluctuate more from day to day and fluctuate more consistently over all conditions than do scores with the numbers. Thus on Day 2 higher average scores were obtained under all conditions with syllables than on Day 1, and scores on Day 3 were all lower than on Day 2. This effect may have been produced by greater variations in the difficulty of the syllable lists used on successive days than in the difficulty of the number lists. One might expect the response availability of letter bigrams to vary more than the response availability of two-digit numbers, and it is possible that this factor produced greater variations in the difficulty of syllable lists than of number lists. More will be said about this below.

Table 2 shows the average amount of information transmitted per *S* for each of the four conditions and the results of some statistics performed on these scores. Since there was a slight tendency for information scores to have a skewed distribution towards the upper end of the scale, Wilcoxon's (1949) paired replicates method was used to evaluate the significance of differences between conditions. It is apparent that there is a tendency for information transmitted to increase as the number of alternatives from which selection is made is increased. The only exception to this trend is the score for number Recognition 90, and it is possible that this is abnormally low due to chance factors. It is noteworthy that in terms of information transmitted performance was not better on recognition than on recall: for both letters and numbers Recall was significantly better than Recognition 30 and for numbers Recall was significantly better than Recognition 90.

A rank-order correlation coefficient was computed for the correlations between individual *S*'s information

TABLE 2

AVERAGE INFORMATION TRANSMITTED PER *S*

Cond.	Letters			Numbers		
	Average *I* score in bits	Significance of difference between means	*p*	Average *I* score in bits	Significance of difference between means	*p*
Recog. 30	6.2	<.02	.32	5.0	<.05	.11
Recog. 60	9.9	*ns*	.10	9.0	*ns*	.21
Recog. 90	11.9	*ns*	.35	6.6	<.01	.33
Recall	11.5			11.1		

scores under Recognition 30 and 60, Recognition 60 and 90, and Recognition 90 and Recall. Although all correlations are positive none are in fact significant, but the correlation between Recall and Recognition 90 is at least as high as the correlations between different recognition conditions.

Discussion

Application of information theory.—The calculation of amount of information passed per *S* by the method used here depends on a number of assumptions: these assumptions are discussed in detail in Judd and Sutherland (1959). Of the assumptions there made there are two which were probably not completely fulfilled in the present experiment (Assumptions 5 and 7). Assumption 5 states that the a priori probability of a given symbol appearing in the set of symbols received should be equal for all symbols. If the response availability of all symbols is not equal this assumption will not be met. Underwood and Schulz (1960) have shown that the response availability of letter bigrams does vary, and this undoubtedly applies to our syllables: it may also apply though to a lesser degree to the numbers. This means that our information scores are likely to be underestimates of information transmitted, but the relative underestimation is likely to remain constant from condition to condition unless response availability differs from condition to condition which there is no reason to suppose it does. Thus the estimates of relative amount of information transmitted under different conditions are likely to be reasonably accurate. Assumption 7 states that no information shall be conveyed by wrong responses, i.e., that wrong responses are equally distributed over all symbols in the set from which selection is made. No attempt has been made to analyze in detail the present data to discover whether wrong guesses were correlated in any way with the stimulus lists. The design of the experiment does not facilitate this type of analysis since if *Ss* are grouped by both condition and list, there are only six *Ss* in each group. However, random sampling of individual *Ss*' response lists did not reveal any tendency to score higher than chance expectancy on "near misses." Further experiments are in progress which lend themselves better to this type of analysis. Moreover, even if this condition is not satisfied, relative scores under different conditions will not be affected unless there is a difference in the tendency to make near misses under different conditions, though the absolute amount of information transmitted will be underestimated.

The low correlations between individual *Ss*' scores under different conditions are probably partly due to lack of marked differences in ability between *Ss*. In addition, each score used in correlation is based on one trial only. Since the number correct on one trial is partly determined by chance factors, *S*'s performance on any one trial is not necessarily a correct reflection of her ability. For example suppose that over a series of trials with $k = 90, q = g = 15$, *S* would have averaged 9 correct responses per trial. On the assumption that on individual trials the number correct will be randomly distributed with the constraint that the average correct will tend to 9 over a large series of trials, this *S* will get 12 correct once in about 30 trials. In this event, the information score would be taken as 28.8 bits, whereas if the true average of r over a large series of trials is only 9, the correct estimate of information transmitted is only 13.5 bits (cf. Fig. 1 of Judd & Sutherland, 1959). These chance fluctuations in the value of r should of course cancel out over a large number of *Ss*.

The result that information transmitted tends to increase as the number of alternatives from which selection is made increases is in line with other findings applying information theory to memory: thus Miller (1956) has found that in immediate memory the

amount of information transmitted increases as the size of the vocabulary from which selection is made is increased. The present results suggest that Miller's finding may be generalized to memory situations other than that of immediate memory. This finding is particularly striking in view of the fact that *S*s take longer to finish checking numbers in Recognition 90 than in Recognition 30. In addition we might expect some retroactive inhibition effect from the symbols in the recognition list itself to operate on the final responses made. This effect would presumably be greater in the case of the 90 list than in the case of the 30 list, yet in terms of information transmitted *S*s do better in recognizing out of 90 than in recognizing out of 30.

Recognition and recall.—There are three difficulties in making direct comparisons between recognition and recall: (*a*) As already pointed out retroactive inhibition from the recognition list itself may lower performance in recognition by adversely affecting later responses in recognition. (*b*) In recall *S*s tend to make the first few responses faster than in recognition but to make later responses at a slower rate than in recognition. (*c*) If there is a difference in the psychological mechanism at work in recognition and recall, this difference may only show itself in performance if individual *S*s adopt the correct set for each of the two conditions. It must be remembered that these difficulties have existed in previous experiments, although they have not always been noted.

Despite these difficulties, the hypothesis to be tested, namely, that recognition is superior to recall only because it usually involves selection from fewer alternatives, is confirmed by the present findings. It is particularly striking that in terms of information transmitted Recall is much superior to Recognition 30. These two conditions approximate most closely to the conditions used in previous experiments since recognition lists which are equal to or nearly equal to the total number of alternatives from which selection is made in recall have not previously been used.

The difficulty of retroactive inhibition affecting recognition may go some way towards explaining why Recall of numbers was significantly better than Recognition 90, but it is not obvious why Recall of syllables was not also better than Recognition 90. Since each number condition always followed the equivalent letter condition this may indicate that practice on letter recall helped *S*s to adopt a recall method for the number recall and thus improved performance. It will be remembered that in the original experimental design the syllable lists were included in order to give *S*s practice on each condition before doing the task with numbers. Because each letter condition always preceded the same number condition, it is not worth speculating further on the *differences* between the results for letters and numbers. Although an effort was made to control set in the instructions, some *S*s during recognition clearly *recalled* the first few numbers checked since they would sometimes check numbers scattered randomly through the recognition list before resorting to scanning the list. The fact that significantly more responses were given in recognition out of 90 than in recall suggests that there is some difference between the psychological processes involved in the two conditions.

In conclusion it may be worth saying something about the generality of this approach to problems of recognition and recall. The main difficulty lies in ensuring that Assumptions 5 and 7 in Judd and Sutherland (1959) are met. It should be emphasized that this difficulty is a practical difficulty rather than a difficulty of principle. Equation 3 in that paper gives a formula for computing information conveyed in recognition and recall tasks taking into account both unequal a priori probabilities in the response set and also the probabilities of given wrong symbols appearing in the response set being conditional on which symbols appear in the set sent. Thus this formula would enable us to compute accurately the information transmitted despite varia-

tions in the probabilities of symbols appearing in the response set due to both difference in response availability and the occurrence of "near misses." As explained in the paper referred to, the reason why this formula cannot be applied is that in order to apply it accurately an impossibly large number of experiments would need to be performed. It may be that the formula used in the present paper to compute information scores could be adapted to take into account differences in response availability and "near misses," but this would require considerable mathematical work and the final formula would be very much more complicated than that used here. The advantage of Equation 24 is that it enables us to compare performance over a variety of different conditions by means of a rational measure; its disadvantage is that its use requires that response availabilities and the proportion of near misses shall not change from condition to condition.

It is also difficult to apply the formula in every day life situations involving recall and recognition. We normally suppose that it is much easier to recognize a human face than to recall one. If by recognizing a face we merely mean that we realize we have seen it before and by recalling we mean that we are able to recall accurately the right face given the stimulus of the person's name, then we are clearly transmitting much less information in the former case than in the latter. If we assume that there are n possible faces, of which we have seen r, that all faces have an equal probability of occurrence, and that we are always correct, then in the former case we transmit $-\log_2 r/n$ bits of information, in the latter $-\log_2 1/n$ bits: since r is much larger than 1 we clearly convey much less information in the recognition case than in the recall case: we would therefore expect recognition to be easier than recall. Sometimes we use recognizing to mean identifying: in this use of the word it is not obvious that we do in fact do better at recognizing faces than at recalling them. Even here however, we might expect better recognition than recall on the present analysis, since we shall say that someone is correctly identified if any identifying mark can be provided. Whereas in recalling a face, to perform correctly someone must recall the correct face out of all possible faces, a face will be correctly identified if any one of a number of correct identifying characteristics is provided: e.g., if the correct name out of all possible names is selected, if the place in which the face was last seen is correctly selected out of all possible places, if the occupation of the owner of the face is specified, if his address is specified, etc. Thus in recalling faces S is forced to select one alternative out of one ensemble, but in identifying faces S can usually select one alternative from each of a number of alternative ensembles. If the ensemble he is to choose from is not specified in advance this is again clearly an easier task, and in addition some of the ensembles may be smaller than the number of possible faces.

This analysis of recognition and recall in everyday life is extremely crude, but is intended only to show how the ideas behind the present experiment might be applied to explaining these phenomena in everyday life. Unfortunately, because of the impossibility of setting limits to the stimulus and response ensembles, it is not possible to test the present hypothesis in the everyday life situation. This limitation, however, applies equally to most psychological hypotheses, and it is precisely for this reason that hypotheses have to be tested by using simple standardized material in the laboratory, and this is what we have attempted to do in the experiment described above.

Summary

In previous experiments on recognition and recall, there has been no clear limit to the number of alternatives from which *S*s were selecting in recall: the superiority of recall over recognition normally found may be due to selection being made from a larger ensemble

in recall than in recognition. The present experiment was designed to test the hypothesis that the amount of information conveyed in recognition is the same as that in recall. The *S*s were given one presentation of lists of 15 two-digit numbers and 15 two-letter syllables and then asked to recall or recognize symbols from the list presented. Each *S* served in four conditions: recognition out of a list of 30, recognition out of 60, recognition out of 90, and recall from an ensemble of 90. Performance scores were transformed into a measure of the amount of information transmitted per *S* under each condition. It was found that at least as much information was transmitted in recall as in recognition and that the amount of information transmitted in the recognition conditions tended to increase as the number of alternatives from which selection was made increased.

References

JUDD, B. R., & SUTHERLAND, N. S. The information content of nonsequential messages. *Inform. Control*, 1959, **2**, 315–332.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, **61**, 511–519.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.

(Received April 5, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 430–436

PERCENTAGE ESTIMATION OF PROPORTION AS A FUNCTION OF ELEMENT TYPE, EXPOSURE TIME, AND TASK [1]

EMIR H. SHUFORD

University of North Carolina

The scale of proportion, based upon the responses of *Ss* to randomly composed arrays of two or more types of elements, is a perceptual task of some interest (Stevens & Galanter, 1957). Furthermore, an understanding of the processes involved in this task is necessary to the use of such stimuli for the control of subjective probability, e.g., see Shuford (1959).

In experiments conducted by Philip (1947) and by Stevens and Galanter (1957) the stimuli were cards containing random arrangements of 36 colored dots, typically of two different colors. Philip (1947) had *Ss* respond with an 11-point scale to cards containing from 13 to 23 (in steps of 1) dots of a specified color. The relation between average category number and proportion was linear. Stevens and Galanter (1957), using cards containing from 3 to 33 (in steps of 3) dots of a specified color, had *Ss* respond with a 7-point scale and with a percentage estimate of the proportion of dots of a given color. In this case the relation between average response and proportion was an inverse ogive with the curvilinearity more extreme for the 7-point scale.

It appears likely that this inverse ogival relation is due to the fact that *Ss* can perfectly discriminate the more extreme proportions when so few dots are on the cards. For example, *Ss* could count or subitize the 3 or 6 dots of the less frequent color. Therefore, the first category could be used for 3 dots, the second for 6, the sixth for 30 (i.e., 6 dots of the other color) and the seventh for 33, leaving three categories for the remaining seven proportions. As for the percentage estimates, *Ss* could figure that $3/36 = 1/12$ is about 8% and that $33/36 = 1 - 1/12$ is about 92%. Thus there would be almost no error in the most extreme proportions. However, *Ss* might be uncertain as to the exact values for the intermediate proportions since it is more difficult to count these larger numbers. Shuford and Hall (1959) and Shuford and Wiesen (1959), in re-interpreting some of the phenomena of psychophysical judgment, have shown that under this condition the optimal decision-theoretic strategy is to regress the estimates toward the center of the stimulus distribution. The amount of regression is a function both of the degree of uncertainty about the true stimulus value and of the distance between the stimulus and the center of the stimulus distribution. Therefore, it is not unreasonable to expect an inverse ogival relation when the more extreme proportions can be counted by *Ss*.

The use of stimuli composed of a much larger number of elements should reduce the curvilinearity since the number of elements required for a given proportion is much larger than in the case of 36 elements. It should be noted that it is impossible to eliminate this tendency completely. If *Ss* are shown stimuli composed of only one type of element they invariably respond with 0% or 100%. Thus, regardless of the total number of elements, the relation must go through (0, 0) and (100, 100). Increasing the total number of elements serves only to decrease the range of extreme proportions for which the scale of proportion is confounded by the possibility of *Ss* counting the

[1] This research was supported in part by ARDC Project No. 7740 of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas, and by a United States Air Force Research Contract AF-49(638)-729 with the Air Force Office of Scientific Research, Air Research and Development Command. The author expresses his appreciation for the advice of Fred Attneave, Ward Edwards, and Tom Jeffrey and for the assistance of Earl Lingle.

number of elements of the less frequent type.

Stimuli chosen for the present experiments were 20 × 20 matrices (see Fig. 1) which contained varying proportions of elements of two types randomly arranged. The smallest number of elements of a given type was 40, a number too large to be counted or subitized by *Ss* even during long exposure times. Therefore, *Ss* cannot know the actual proportion with certainty and must base their response either on a sample or on a vague impression of the matrix. The experiments provide some information concerning the processes by which *Ss* derive percentage estimates of proportions.

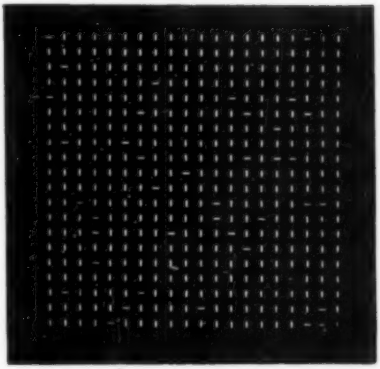


FIG. 1. A random matrix composed of 90% vertical bars and 10% horizontal bars.

METHOD

Subjects.—The *Ss* were 636 basic airmen in training at Lackland Air Force Base who were taken from a subject pool with no attempt at further selection.

Apparatus.—The apparatus consisted of an automatic slide projector, a large projection screen, 35- mm. slides (40 in a magazine), and answer sheets. The slides were either black and white negatives or color positives of the matrices. Each black and white slide contained a matrix of vertical and horizontal white bars on a black ground. Each color slide contained a matrix of red and blue squares on a white ground.

Twenty location matrices, two for each of the percentages 10, 20, 30, 40, 60, 70, 80, 90, and four for 50, were obtained by taking the five pairs of percentages (e.g., 10 and 90), determining two random arrangements for each pair, and interchanging the positions of the two types of elements for each randomization. These location matrices provided the random patterns for both the vertical and horizontal bars and the red and blue squares. Each stimulus matrix was photographed twice yielding a set of 40 slides for the matrices composed of red and blue squares and a set of 40 slides for the matrices composed of vertical and horizontal bars.

Procedure.—The *Ss* participated in the experiments in groups of 10 to 16. They were seated in writing chairs arranged so that each *S* had a clear view of the screen. The *Ss* were then shown an original matrix (not a slide) with a proportion of 50% and were told that they would be shown photographs of different matrices like this one. In Exp. I, *Ss* were instructed to estimate the percentage of elements of the type specified by *E*, e.g., the percentage of vertical bars. In Exp. II, *Ss* were instructed to estimate the percentage of elements of that type which is more frequent (or less frequent) in the particular matrix and to indicate which type they were estimating by recording the letters R and B or H and V after their estimate.

Experiment I investigated the effect of exposure time, either 1 or 10 sec., and type of stimulus element, either red and blue squares or vertical and horizontal bars, on the percentage estimation of proportion. These two factors yielded four experimental conditions: RB-1 ($N = 132$), RB-10 ($N = 120$), HV-1 ($N = 132$), and HV-10 ($N = 152$).

In Exp. II, conducted after the completion of Exp. I, only the 10 sec. exposure time but both sets of slides were used. Therefore, there were two experimental conditions: RB-10 ($N = 54$) and HV-10 ($N = 46$).

Once the task had been explained to *Ss*, the lights were dimmed and the appropriate set of 20 matrices (either RB or HV) was presented twice in the same randomly determined order. The *E* called "ready" before each slide was shown; *Ss* looked at the slide for either 1 or 10 sec. and recorded their estimates (any integer from 0 to 100) on answer sheets. The *Ss* were not informed as to the accuracy of their estimates. The exposure time and 8-sec. intertrial interval were controlled by an automatic sequencing device.

After the 40 trials were completed, *Ss* received a brief intermission and then returned and gave the complementary response to each of the *same* set of 40 slides. For example, approximately half of the *Ss* in Cond. RB-1 estimated the percentage of red squares during the first 40 trials and then estimated the percentage of blue squares during the last 40 trials. The other half of the *Ss* in this condition estimated the blue squares during the first 40 trials and the red squares during the last 40 trials. In Exp. II approximately half of the *Ss* in each condition estimated the larger proportions during the first 40 trials and the smaller proportions during the last 40 trials. This order was reversed for the other *Ss*. Since the distribution of proportions is symmetrical for each set of matrices, each type of element was more frequent (or less frequent) on 16 of the 40 trials. Both types were equally frequent on the other eight trials.

The session required up to an hour to complete depending upon the exposure time.

Results

The estimates were averaged over *Ss* for each trial and then examined for any evidence of learning. In general, the differences between mean estimates and proportions tended to be slightly smaller for the later trials.

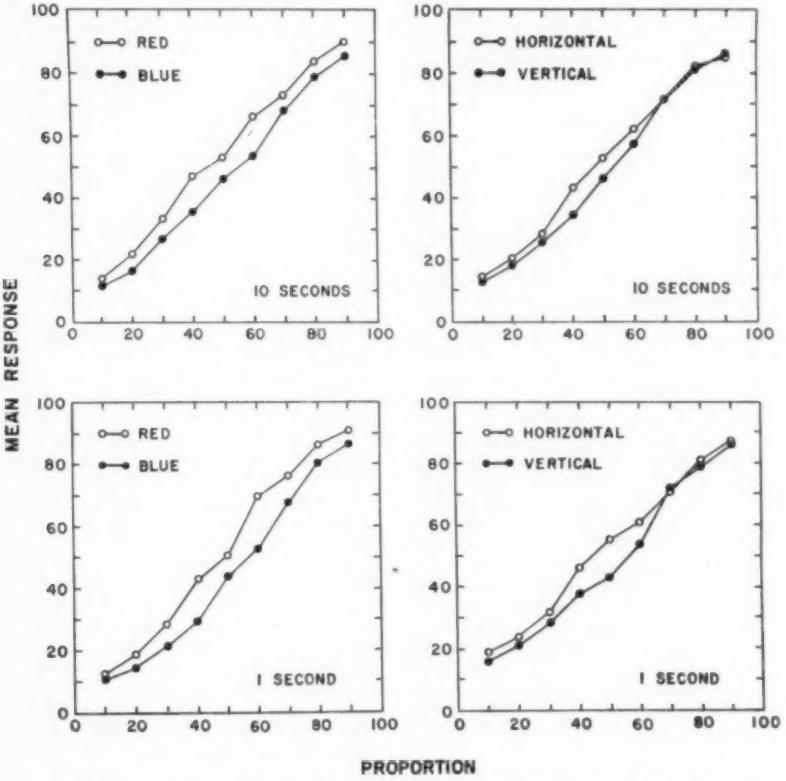


FIG. 2. Mean response vs. proportion as a function of element type, RB or HV, and exposure time, 1 or 10 sec. Data points for 50% are based on $8N$ observations; the other data points on $4N$ observations where $N = 120$ for Group RB-10, $N = 132$ for Group RB-1, $N = 152$ for Group HV-10, and $N = 132$ for Group HV-1.

Since there was no appreciable practice effect and since any differences in responses to matrices containing the same proportion of elements must be attributed to random arrangement of elements in the matrices and to random ordering of the slides, the data were averaged so that mean response could be plotted as a function of proportion for each element type.

Figure 2 shows mean response as a function of proportion for each of the four conditions of Exp. I. In each case the relation between percentage estimate and proportion appears to be not the *inverted* S-shaped curve found by Stevens and Galanter (1957) but, rather, the usual S-shaped ogive rotated slightly about the point (50, 50) in a clockwise direction. Notice that the proportion of red squares and of horizontal bars is overestimated relative to the proportion of blue squares and of vertical bars. This bias is greater for the colored squares than for the bars. The bars were probably better matched for brightness since a horizontal bar was the same size as a vertical bar. Notice, also, that the amount of bias appears to be greater for the shorter exposure time and for porportions close to 50%, suggesting that the amount of bias may be related to the variability of the estimates.

To check this possibility, the variances of the responses were computed for each trial, averaged for matrices with identical proportions, and plotted against proportion. Figure 3 shows average response variance (combined for both vertical and horizontal bars) as a function of proportion for the 1- and 10-sec. exposure times. Notice that the variances for the 1-sec. exposure time condition are consistently above those for the 10-sec. exposure time and that both curves

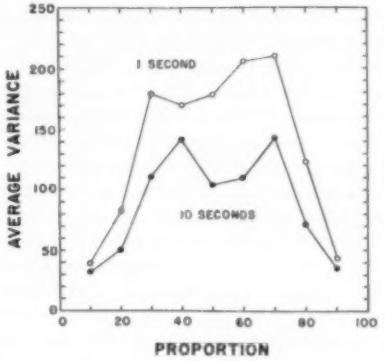


FIG. 3. Average variance vs. proportion for Groups HV-1 and HV-10.

appear to be inverted U-shaped functions similar to the function relating the variance and mean of a binomial distribution. The bimodal tendency of the two curves is in agreement with a previous finding (Stevens & Galanter, 1957, p. 405)—"Actually there is some indication that near the exact center of the range, where the dots of the two colors are equal in number, discrimination improves over what it is on either side of center." The bimodal tendency would be reduced, however, if the variances were plotted against mean response instead of proportion. Also, there was an increased tendency for *S*s to respond with a judgment of 50% for the central stimuli.

Correlations between responses to repeated presentations of the same matrix ranged from about .8 for the more extreme proportions to about .3 or .4 for the intermediate proportions. Thus, there were characteristic differences between *S*s, which were more pronounced for the more extreme proportions.

These correlations and variances can be used to obtain a measure of the reliability of the data points of

Fig. 3. An *SD* of .913 for the data points corresponding to the intermediate proportions (excluding that for 50% which is based on twice the number of observations) for Cond. HV-1 was obtained by using a variance of 200 for each observation, a correlation of .4 between repeated observations, and a sample size of 132. The *SD*s for data points corresponding to more extreme proportions and for other experimental conditions will be about the same or smaller. For example, an *SD* of .507 for the data points corresponding to proportions of 10% and to 90% for Cond. HV-1 was obtained using variances of 40, correlations of .8, and a sample size of 132. The small sizes of these *SD*s and the fact that similar results were obtained for corresponding conditions suggest that the data are highly reliable.

The binomial character of the response distributions was studied further by randomly selecting one trial for each of the nine proportions in one experimental condition, choosing $n = 12$ as the sample size for the set of binomial distributions which would adequately represent the data and calculating χ^2 for the goodness of fit of each empirical frequency distribution to the corresponding binomial distribution with the same mean. The fit was not bad for the intermediate proportions but was very poor for the more extreme proportions, particularly in the tails of the distributions.

One factor (in addition to individual differences in, for example, sample size) contributing to this lack of fit is that *S*s rarely responded with an estimate greater than 95 or less than 5. In particular, they never responded with 100% or 0%. This truncation is probably related to the total number of elements in the matrix and seems to be inherent in the nature of the task.

Another factor is that *S*s sometimes forgot (temporarily) which type of element they were supposed to be estimating. Some *S*s commented after the experiment that this had happened to them. This was also indicated clearly in *S*s' responses to a proportion of 90%. The great majority of estimates typically were

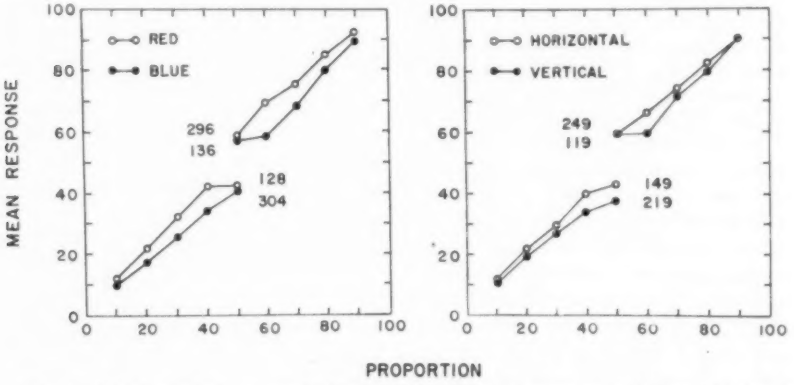


FIG. 4. Mean response vs. proportion as a function of element type, RB or HV, for *S*s estimating larger proportions (upper curves) or smaller proportions (lower curves). Data points for proportions other than 50% are based on $4N$ observations where $N = 54$ for Group RB and $N = 46$ for Group HV. Numbers of observations for data points corresponding to a proportion of 50% are given by the numbers in the graphs.

above 75 while a few fell around 10. The lack of intermediate estimates suggests that the extreme errors were due not to random fluctuations in *S*s' perceptions of the matrices, but rather to reversals in the association of element type and proportion.

Figure 4 shows mean responses as a function of proportion for each of the two conditions in Exp. II. The data points for proportions other than 50% are based on *S*s' estimates for the four slides with the stated proportion of elements of a given type, ignoring *S*s' statements as to the type of element being estimated. This information was, necessarily, taken into account in obtaining the data points for a proportion of 50%. For example, *S*s estimated the red squares 296 times and the blue squares 136 times when asked to estimate the percentage of elements of the more frequent type in the eight slides containing a proportion of 50%. The data points in the upper half of each graph were obtained when *S*s were estimating the larger proportions; those in the lower half, when *S*s were estimating the smaller proportions.

Notice that the S-shaped character of the curves has largely disappeared. In particular, there is almost no tendency to overestimate a proportion of 10% or to underestimate a proportion of 90%. The displacement of the curves away from a response value of 50% is, apparently, a consequence of the instruction to estimate the larger (or smaller) of the two proportions. The red-blue and horizontal-vertical bias is still present, and again, is greater for the colored squares than for the bars. This bias is also reflected in the frequencies with which *S*s classified the element types when both types were equally frequent.

The number of misclassifications, e.g., responding 70R when there were actually more blue than red squares in the matrix, was counted for each proportion other than 50%. The percentage of errors for the 40–60 RB matrices was about 19%; for the 40–60 HV matrices, about 27%. The error percentages for the other proportions were relatively constant at about 3% for RB and 13% for HV and were due primarily to reversal type errors.

Discussion

There is evidence that the perception of proportion has some of the characteristics of a sampling process. For example, the U-shaped relation between response variance and proportion found in Exp. I suggests that *S*s may, in effect, base their estimates on a sample of elements in the matrix. A similar relation was found by Green, Wolf, and White (1959) in an experiment using random matrices which contained bar patterns defined by element probabilities. They obtained probability difference at threshold (for detecting a bar pattern) as a function of average element probability at threshold, p. This relation was proportional to $(pq)^{\frac{1}{2}}$ but larger by a factor of 5.3, suggesting that *S*s were, in effect, using only a portion of the elements in a matrix.

Shuford and Wiesen (1959), using a set of 16×16 random matrices composed of 0's and 1's, studied the effect of short exposure times (1/2, 1/10, and 1/50 sec.) on percentage estimation of proportion. By assuming a theoretical sampling model for *S*'s responses, they were able to obtain an estimate of sample size for each exposure time. The sample sizes of some *S*s, particularly those who had small sample sizes, were unaffected by reducing exposure time. If the sample size for an individual *S* did change, it did so not by a factor of 1/5 or 1/25 (proportional to the reduction in exposure time) but, typically, by a factor of 1/2 or 1/3. These results, though tentative, suggest that *S*s successively fixate on different areas of a matrix, each time observing a relatively constant number of elements.

The ogival nature of the curves obtained in Exp. I might be accounted for by assuming that *S*s observe samples of elements in a matrix and that they tend to fixate on an area in which there is a cluster of elements of both types. If the curvilinearity were due to this biased sampling process then it also should be found in the curves for Exp. II. There is little evidence for this in Fig. 4.

It should be mentioned that when Toda's game is used to obtain subjective probabilities based on randomly selecting an element from one of these matrices, most of the individual relations between subjective probability and proportion are linear, with slopes less than 1 and intercepts greater than 0 (Shuford, 1959).

These findings suggest that the S-shaped relations observed in Exp. I are due not to a perceptual bias, but rather, to a response bias, e.g., the truncation of extreme estimates close to 0% or 100% observed in the response distributions.

On the other hand, a biased-sampling interpretation seems to account very well for the red-blue and horizontal-vertical bias found in both experiments. Assume, again, that *S*s fixate several times and observe a sample of elements in a matrix. Assume, further, that *S*s tend to fixate on areas composed of clusters of red squares or horizontal bars. Thus, the samples perceived by *S*s are biased so as to yield too many red squares and horizontal bars. Furthermore, the amount of bias is a function of the variance of the sample proportions. The size of this variance depends, of course, upon both exposure time and proportion.

The validity of any such interpretation can be determined only by further experimentation. This sampling theory suggests that there are discontinuities in the relation between response variability and exposure time. A more direct test is to photograph *S*'s eye movements and observe the relation between response error and number of fixations.

As to the scale of proportion, it is evident that if this is defined as the relation between a response measure and proportion, there are many such scales depending upon the response measure, element type, exposure time, task, individual *S*, etc. However, one might find it useful to think of the *scale of proportion* as an intervening variable depending only upon proportion. In this sense, then, the author prefers to think of the *scale of proportion* as the identity relation.

Summary

Two experiments were conducted in order to investigate the effect of element type, exposure time, and task on the percentage estimation of proportion. The stimuli were photographs of 20 × 20 matrices made up of randomly arranged red and blue squares or vertical and horizontal bars.

When *S*s estimated the percentage of elements of the type specified by *E*, they overestimated the proportion of red squares and of horizontal bars relative to the proportion of blue squares and vertical bars. The amount of this bias was smaller for the longer exposure time and for the more extreme proportions. The observed relations between mean response and proportion were S-shaped.

When *S*s estimated the percentage of elements of the more frequent (or less frequent) type, the proportion of red squares and of horizontal bars was again overestimated, but the ogival nature of the relation between mean response and proportion was absent.

The results are interpreted in terms of a hypothetical sampling process and various response biases.

References

Green, B. F., Jr., Wolf, A. K., & White, B. W. The detection of statistically defined patterns in a matrix of dots. *Amer. J. Psychol.*, 1959, **72**, 503–520.

Philip, B. R. Generalization and central tendency in the discrimination of a series of stimuli. *Canad. J. Psychol.*, 1947, **1**, 196–204.

Shuford, E. H. A comparison of subjective probabilities for elementary and compound events. *U. N.C. Psychom. Lab. Rep.*, 1959, No. 20.

Shuford, E. H., & Hall, W. J. A decision theory approach to psychophysics. *U. N.C. Psychom. Lab. Rep.*, 1959, No. 22.

Shuford, E. H., & Wiesen, R. A. Bayes estimation of proportion: The effect of stimulus distribution and exposure time. *U. N.C. Psychom. Lab. Rep.*, 1959, No. 23.

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received April 9, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 437–441

CONCEPT IDENTIFICATION AS A FUNCTION OF LANGUAGE PRETRAINING AND TASK COMPLEXITY

ELIZABETH A. RASMUSSEN AND E. JAMES ARCHER

University of Wisconsin

In concept identification experiments, *S* is typically given a number of stimuli and has to organize them into categories on the basis of some common property or properties. The concept may then be defined as the class name or label for the grouping of the stimuli. The main purpose of this type of research is to observe *S*'s behavior as he tries to find the label for the class. Most of the work in concept identification, up to this time, has investigated the effect of such variables as task complexity (e.g., Archer, Bourne, & Brown, 1955), nature of the concept to be attained (e.g., Heidbreder, 1946), and previous experience (e.g., Kendler & D'Amato, 1955). The present study was designed to examine the effect of teaching *S* some of the labels before he identified the concepts.

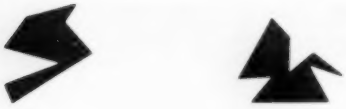
From a review of the work done on these variables, it appears that fairly consistent results have been obtained on the effect of task complexity, conflicting results prevail in the work done on the nature of the concept, and that only limited investigation has been made on the effect of previous experience. In regard to the last two of these variables, it seems logical to hypothesize that the amount of previous experience that *S* has had with the stimulus objects would influence the manner in which he categorized the objects. For example, Heidbreder's (1946) order of dominance might be explained by this hypothesis. Most *S*s have had more experience with concrete objects and are more likely to have already attached a label to the class of such objects as compared to concepts of abstract forms.

Another potential relationship between prior experience and language appears in the linguistic relativity hypothesis advanced by Sapir and Whorf (Whorf, 1956) which states that the structure of a culture's language influences the manner in which the individual understands reality and behaves with respect to it. The validity of this hypothesis has thus far not been sufficiently tested. This hypothesis would predict that the concepts most readily formed would be those for which our culture has labels and *S*s have had experience in using these labels.

The major purpose of the present experiment was to attempt to control the past experience, through language training, of *S* to some aspect of the stimulus objects, and to see what effect this had on subsequent concept identification performance in which the relevance of the labeled dimension and the complexity of the task were varied. The *shape* of nonsense forms which varied in several dimensions was used as the dimension to be labeled as this appears to be the only available property of visual stimuli for which *S* does not already have a name.

PROCEDURE

Subjects.—The *S*s in this experiment were 128 students (64 men and 64 women) who were enrolled in introductory psychology courses at the University of Wisconsin and



LATUK TAROP

FIG. 1. The two random shapes which were previewed and classified by all *Ss*. The names of *Latuk* and *Tarop* were learned by the language pretraining *Ss* before they attempted the concept identification tasks.

who received experimental points for their service. The only restriction was that they could not have served in a similar experiment previously. Each *S* served individually for 10 to 40 min. depending on his ability to solve the task.

Materials.—The stimuli used consisted of two nonsense shapes constructed following Method 1 of Attneave and Arnoult (1956). The shapes were angular, had closed contours, and were of approximately the same area. The stimuli were varied in five bilevel dimensions: (*a*) shape, 8- or 10-sided; (*b*) size, large (.543 sq. in.) or small (.155 sq. in.); (*c*) color, yellow or blue; (*d*) number, one or two identical stimulus figures; and (*e*) horizontal position, figure(s) appeared on the left or right side of the screen.

Design.—A 2^5 factorial design was used in this experiment. The sex of *S* was one variable; all other variables were manipulated by *E*. The independent variables were: (*a*) sex; (*b*) type of pretraining, language or aesthetic; (*c*) degree of pretraining, low or high; (*d*) amount of irrelevant information in the concept identification task, 1 or 3 bits; and (*e*) relevance of the labeled dimension, shape relevant or shape irrelevant. All possible combinations of these variables were used, making a total of 32 different conditions. Four *Ss* served in each condition.

Pretraining.—For the pretraining task, *S* was assigned to either a language or aesthetic group and received either a low or high degree of pretraining. In the language group *S* was presented with a paired-associate learning task in which he had to associate one of two paralogs from Noble's list (1952) with one or the other of the two nonsense shapes. The two figures and their assigned names are shown in Fig. 1. Four of the stimulus dimensions were used in this task but only shape was relevant; color, size, and number were irrelevant to the learning situation. Each stimulus pattern (e.g., two large blue 10-sided nonsense shapes) was first presented by itself and then appeared again with its "name" beneath it. Each frame appeared for 2 sec. and *S* was instructed to anticipate the correct response before the name appeared with the nonsense figure. The stimuli were presented on a strip film in five different random orders of the 16 possible stimulus patterns. The *Ss* learned the associations to a low or high degree which consisted of either one or all five orders of presentation of the 16 stimulus patterns. The *S* spoke his responses aloud and they were recorded by *E*. The number of presentations for each degree of learning had been determined by a pilot study conducted prior to the beginning of this experiment.

The *Ss* in the aesthetic pretraining group were presented with the same stimulus patterns in the same order as those in the language group. The stimuli were again presented in a paired-associate fashion, except instead of a name appearing with the second presentation of the stimulus, a linear scale with the numbers from 0 to 100 appeared. In this condition *S* was told to judge the "pleasantness" of the stimulus pattern by assigning it a number between 0 and 100. Zero indicated the most unpleasant pattern and 100 denoted the most pleasant one. With this method these *Ss* had as much familiarity with the stimuli as did the *Ss* in the language group, but no particular dimension was emphasized.

Concept identification task.—Upon completion of the pretraining session, *S* was told that he would see a series of patterns similar to those he had just seen and that he now would have to classify the patterns into four categories. The *S* indicated his choice of category for each stimulus pattern by pressing one of four unlabeled switches on a control panel; each switch represented a combination of two levels of two relevant dimensions, e.g., (*a*) yellow and 8-sided; (*b*) yellow and 10-sided; (*c*) blue and 8-sided; and (*d*) blue and 10-sided. A dimension is relevant if it is necessary for a correct classification of the patterns. After *S* had indicated his choice by pressing one of the switches, the correct response, whether or not *S* had been correct, was immediately indicated by a light above the switch which would have been correct. After 2 sec. the information light turned "off" and 1 sec. later a new pattern was presented.

Two of the stimulus dimensions were always relevant to the task. When *shape* was relevant, *color* was also relevant. When

shape was irrelevant, *size* and *color* were relevant. One or three irrelevant dimensions also appeared with the relevant dimensions. When there was only one irrelevant dimension it was either *size* or *shape* depending upon what was not included in the relevant information. If three dimensions were irrelevant, then either *size* or *shape* was irrelevant plus *number* and *horizontal position*.

Apparatus.—The *S* viewed the stimuli at a distance of about 2 ft. on a 7½ by 11½ in. milk glass screen positioned at eye level in an opaque partition which separated *S* from *E* and *E*'s apparatus.

The stimulus patterns were on strip film and were presented by a Dunning Animatic strip-film projector located on *E*'s side of the screen.

The time to criterion was measured with a Standard Electric clock. The clock was started with *S*'s first response and was stopped when *S* reached the criterion of 32 consecutively correct responses. The responses were manually recorded by *E*.

Results and Discussion

Comparability of groups.—The statistical analyses used in this study assume that the groups of *S*s for each condition were initially equal in problem-solving ability. Since *S*s were assigned randomly to the experimentally manipulated conditions, it is believed that this assumption is tenable.

Response measures.—In addition to time-to-criterion and errors, a ratio of errors per minute (a measure of inefficiency) was determined for each *S*. These measures, of course, are significantly correlated. The product-moment correlation between time and errors was .92; between errors per minute and time was .63; and between errors and errors per minute the correlation was .82. An analysis of variance was used to determine the significance of each of the five independent variables and their interactions. The means for each of the levels of the variables are presented in Table 1. Pairs of means which differ significantly ($P < .01$) are shown in italics. There were 1 and 96 *df* for each of the *F* ratios.

Rather than examine differences

TABLE 1

Means of Time to Criterion, Incorrect Responses, and Errors per Minute for the Major Variables

Variable	Level	Mean Time to Criterion (Min.)[a]	Mean Number of Incorrect Responses[b]	Mean Errors per Minute[c]
Type of pretraining	Language	10.72	*43.09*	*3.54*
	Aesthetic	9.01	*27.84*	*2.64*
			(8.06)*	(14.45)
Degree of pretraining	High	10.16	36.59	3.11
	Low	9.57	34.34	3.06
Irrelevant information	1 Bit	*8.46*	*27.72*	2.82
	3 Bits	*11.27*	*43.22*	3.36
		(8.36)	(8.32)	(5.33)
Problems	Shape irrelevant	*8.39*	31.55	3.10
	Shape relevant	*11.35*	39.39	3.08
		(9.24)		
Sex	Men	10.58	38.56	3.08
	Women	9.15	32.38	3.10

[a] The σ_m, based upon the best estimate of error variance for the entire set (30.30), is 0.694.
[b] The σ_m, based upon the best estimate of error variance for the entire set (923.50), is 3.829.
[c] The σ_m, based upon the best estimate of error variance for the entire set (1.766), is 0.167.
* *F* ratios are shown in parentheses and are significant at .01 level when the two associated means are in italics. *F* ratio is significant at .05 level when means are not in italics. All tests had 1 and 96 *df*.

among independent variables within a response measure, it seems more profitable to examine the independent variables across the response measures. In this way, inconsistencies can be considered.

Type of pretraining.—Analysis of the effect of this variable was one of the main reasons for this study. Clearly, it is a significant variable but not in the predicted direction. Except for the time-to-criterion measure (and even this response shows the effect in the same direction) aesthetic pretraining led to better performance than the language pretraining. It is hard to believe that the language pretraining significantly interfered with the subsequent concept identification. It seems more likely that the intended "control" (aesthetic) condition actually aided subsequent performance by encouraging *S* to attend to and discriminate among the several dimensions of the stimulus figures. On the other hand, the language pretraining groups were reinforced for ignoring the other dimensions and for only responding to the dimension of shape. If this post hoc explanation is correct, the use of an "aesthetic judgment" preview might be usefully employed to reduce response variability in future concept identification studies. The preview could serve to familiarize *S*s with the various dimensions of stimulus variation without calling attention to any one in particular.

In addition to the main effect being significant, it was predicted that the interaction between type of pretraining and type of problem (shape relevant or shape irrelevant) should be significant. The prediction was supported only in the errors/minute analysis ($F = 4.676, df = 1/96, P < .05$). If *S*s learned a name for the nonsense figures in the pretraining stage, they solved the problem somewhat more efficiently if form was relevant in the concept identification stage, and less efficiently if form was irrelevant. The opposite was true if they had not learned a verbal label, i.e., had had aesthetic pretraining.

Degree of pretraining.—This variable was without effect.

Irrelevant information.—This variable was included primarily as a check on the sensitivity of the experiment and the expected outcomes obtained; poorer performance was associated with the greater degree of irrelevant information.

Problems.—The study of this main effect and especially the interaction of this variable with type of pretraining was basic to the rationale of this study. However, only in terms of the time-to-criterion measure was there evidence of a significant effect of this variable. When shape was irrelevant the problems were solved more rapidly. Apparently the identification of concepts involving nonsense shapes are more difficult than concepts not involving such information. Though *S*s made more errors when shape was relevant, this effect was not significant.

Sex.—The sex of the *S* was unrelated to task performance.

In concept identification experiments it is usually assumed that *S* can discriminate between levels and dimensions of the stimulus objects which are presented to him; otherwise *S* would never be able to solve the task. Experiments by Whorf (1940) and Brown and Lenneberg (1958) have postulated that the process of categorization and discrimination is largely a function of the language that our culture uses—our language provides us with the categories into which we place objects and also gives us labels to distinguish among objects. Although the present study does not provide experimental support for this

hypothesis, it may be because the nonsense stimuli were too easily discriminated. Perhaps, if the stimuli were less discriminable initially, the enhancement of discrimination as a result of language pretraining could be detected.

Summary

An experiment was conducted to test the effect of language pretraining and task complexity on a concept identification task. A total of 128 *Ss* served in a 2^5 factorial experiment replicated four times. The five independent variables were: type of pretraining, degree of pretraining, amount of irrelevant information, type of problem to be solved, and sex of *S*. In the language pretraining group *Ss* learned names for nonsense shapes which might be relevant to the solution of the concept identification task; in the aesthetic pretraining group *Ss* were asked to judge the stimuli according to their "pleasantness." In the concept identification task 1 or 3 bits of information were irrelevant, and either *shape* and *color* (problem of shape relevant) or *size* and *color* (problem of shape irrelevant) were the relevant dimensions. The *Ss* classified the patterns into four categories to a criterion of 32 consecutively correct responses.

The essential results of the experiment are as follows: (*a*) An increase in the amount of irrelevant information presented in the concept identification task caused an increase in the number of errors made prior to solution and time to achieve the criterion. (*b*) The type of problem to be solved indicated a differential difficulty as measured by time to criterion with the problem of shape relevant being more difficult than the problem of shape irrelevant. (*c*) The aesthetic pretraining group performed better on the concept task than the language group regardless of the type of problem. (*d*) A predicted result occurred in the significant interaction between type of pretraining and problems in the errors per min. analysis with language pretraining giving rise to better performance on the problem of shape relevant than on the problem of shape irrelevant.

References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49,** 153–164.

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53,** 452–471.

Brown, R. W., & Lenneberg, E. H. Studies in linguistic relativity. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, 1958. Pp. 9–17.

Heidbreder, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35,** 191–223.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49,** 165–174.

Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59,** 421–430.

Whorf, B. L. Science and linguistics. *Technol. Rev.*, 1940, **44,** 247–248.

Whorf, B. L. *Language, thought and reality.* New York: Wiley, 1956.

(Received April 14, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 5, 442–448

INFERENTIAL BEHAVIOR IN CHILDREN: II. THE INFLUENCE OF ORDER OF PRESENTATION[1]

TRACY S. KENDLER AND HOWARD H. KENDLER

Barnard College *New York University*

In previous research it has been established that, given a rather simple problem, some children between 3 and 5 yr. of age are capable of combining previously acquired discrete experiences (habit segments) to reach a solution (Kendler & Kendler, 1956). Such solutions are, like learning, subject to the influence of motivation and reinforcement (Kendler, Kendler, Pliskoff, & D'Amato, 1958). The present research was designed to reveal the influence that the order in which the children received the discrete experiences had upon inferential behavior.

This variable was studied for both practical and theoretical reasons. Practically, because it would be useful for future research to know which of the various possible orders of presentation to employ. Theoretically, because certain formulations could be extended to apply to the present problem. Hull's attention (1935, 1952) was drawn to inference-type behavior by Maier's report that rats were capable of inferring (1929). He was so impressed with the necessity for a behavior theory to encompass such results that he said, "Something of the theoretical urgency of this problem is brought home to us when we consider the blind chaos which would result if, in problem situations, premises or habit segments should appear by pairs as if drawn by chance from a huge urn from the supply possessed by the organism. . . . Chance is evidently an element in intelligent behavior but, clearly, the dice of chance must be loaded in some way or problems would never get solved. In short, any adequate theory of higher adaptive behavior must show how the dice are loaded, i.e., how the characteristics of the problem situation are able to evoke the particular combination of acts which alone will serve to extricate the organism from its difficulty" (Hull, 1935, p. 221).

Hull's attempt to account for inferential behavior within the framework of his learning theory was based upon an experimental paradigm which has been adapted in the following manner to both the present and previously cited research on inferential behavior in children. The *S*s are trained on three separate behavior segments, each of which presents a distinctive stimulus, requires a different response, and yields a different reward. Two of these segments, designated as A-B and X-Y, lead to subgoals. The third segment, B-G, leads to the major goal. After being trained on each of these segments

[1] This research was sponsored by a grant from the National Science Foundation. The data were collected by Doris Wells and Naomi Portnoy. The cooperation of the following schools is gratefully acknowledged: Astoria Child Care Center, Bayside Country Day School, Bethlehem Day Nursery, Carousel School, Church of All Nations Nursery School, First Presbyterian Church Nursery School, Forest Park Nursery and Kindergarten, Forest Hills Jewish Center Nursery School, Garden City Cooperative Nursery School, Greenwich House, Lillian Wald Nursery, Our Lady of Loreto Day Nursery, Park Nursery School, Roslyn Nursery School, Stuyvesant Nursery School, and Virginia Day Nursery, Work and Play School.

individually, *S* is presented with a test trial in which only the A-B and X-Y segments are available to him and he is instructed to get the major goal. The solution to this problem is to link the behavior segments by responding to A to acquire B, and then use B to get G. The X-Y segment serves as a control against which to assess the inferential behavior.

In broad outline, the explanation proposed by Hull[2] is that during the acquisition of each segment of behavior, not only the overt S-R connection is strengthened, but *S* also acquires an anticipatory goal response (r_g) appropriate to the goal object. This r_g works backward until stimuli that mark the beginning of the segment become capable of evoking it. For example, in this experiment the major goal (G) is candy. The *S* acquires r_g's appropriate to candy: sucking movements, etc. In due time B, which in this experiment is either a cube or a marble, comes to evoke r_g. When the subgoal segment A-B is acquired, r_g continues to move forward until it becomes associated with A. The stimulus properties associated with the distinctive r_g for the major goal thus become connected with the A-B segment; they do not become associated with the X-Y segment. Since more S-R connections are associated with the response to A, and since these response tendencies summate, the *S* should, when presented with a choice of A or X, and motivated for G, choose the former. Once A is responded to, it produces B, which in turn leads to the response necessary to produce G, by virtue of the previous training. Thus the habit segments are presumed to become linked to produce inferential behavior.

It is characteristic of Hull's explanations to generate many deductions, several of which he made explicit. Here is one, however, that he did not spell out, namely, that inference should occur more readily when the order of training consists of presenting the major goal segment (B-G) before the A-B segment. This order maximizes the conditions for the association between A and the r_g of the major goal.

This is also one of those occasions when Skinner's formulation (1938) engenders a prediction. The assembly of habit segments can be viewed as an exercise in chaining. In setting up a chain of behavior it is usually most efficient to start with the last link, in this case B-G. In that way the discriminative stimulus (B), through its association with the goal stimulus (G), acquires secondary reinforcing powers. This secondary reinforcer can then serve to strengthen the A-B link. Since Y acquires no such additional secondary reinforcing capacity, the X-Y link should not be able to compete successfully with the A-B link. This analysis should engender the same prediction as was derived from Hull with reference to the order of presentation, namely, that the optimum order would be that in which B-G precedes A-B.

An incidental result of previous research (Kendler & Kendler, 1956) had failed to confirm this prediction, but the test was not sufficiently controlled to be conclusive. The present experiment, which is concerned specifically with order, pro-

[2] Deutsch (1956) justifiably criticizes the logical consistency of Hull's analysis of inferential behavior. His criticism would not apply to the present problem because A and X are physically distinct and spatially separate. It might also be mentioned that the difficulty raised by Deutsch would be avoided by emphasizing the associations between distinctive external stimuli and anticipatory goal responses.

vides a more adequate test both of the theoretical formulations and of the role played by the order of presentation on inferential behavior in children.

METHOD

Experimental design.—All of the possible orders of the three habit segments were included, namely: (I) B-G, A-B, X-Y; (II) X-Y, B-G, A-B; (III) B-G, X-Y, A-B; (IV) A-B, B-G, X-Y; (V) A-B, X-Y, B-G; and (VI) X-Y, A-B, B-G. The experimental design consisted of six experimental groups, 32 *S*s to a group, which corresponded to the six orders of presentation.

Subjects.—The *S*s were 212 preschool children, 20 of whom were eliminated due to failure to acquire the separate habit segments. The data to be reported are based on the remaining 192 children (100 males and 92 females) whose ages ranged from 30 to 65 mo. They were randomly assigned to each experimental group without regard to age or sex.

Apparatus.—The series of experiments on inferential behavior in children has involved a continuous search for improved procedures and apparatus. For this reason each of the published reports cited above, and the present research has utilized a different apparatus. While there is a loss in continuity as a result of such changes, it is hoped that the loss will be offset by a reduction in possible artifacts and an increase in flexibility and control.

The present experiment, which provided a more exacting test of inference, utilized a portable aluminum apparatus in which the delivery of the reinforcements was electronically controlled. It consisted of three distinct panels 7 in. high by 6.25 in. wide. Each panel, which could be exposed to *S*'s view singly or in combination with the others, corresponded to one habit segment. The center panel, which was painted yellow, provided for the B-G segment. On its surface, near the top, was a circular opening 1 in. in diameter, into which *S* could drop the small objects, a brass cube or a glass marble, that served as subgoals in this experiment. If *S* dropped the correct subgoal into the circular opening, a small, hard-coated, chocolate-filled candy, commercially known as an M & M, was propelled to an open tray near the bottom of the panel. The candy served as G, the major goal. If the incorrect subgoal were dropped into the aperture, no reward was delivered.

The two side panels corresponded to the A-B and X-Y segments. The left panel was painted red and the right panel was painted blue. Each of these panels was equipped with two manipulanda. Each manipulandum was shaped differently and required a different characteristic movement: "turn" and "toggle" on the left and "push in" and "push down" on the right. The manipulanda on the left panel were arranged on a horizontal axis such that one was to the right of the other. The manipulanda on the right panel were arranged on a vertical axis such that one was above the other. When *S* responded appropriately to the correct manipulandum on either side panel, he closed a circuit which led to the delivery of the appropriate subgoal. When *S* responded incorrectly, no subgoal was delivered.

Each of the three panels could be hidden from the view of *S* by sliding an aluminum door in front of it. Except for the panels the entire apparatus, including the sliding doors, was unpainted.

Procedure.—The data were collected in the nursery school or day care center that *S* attended. In each school a room was assigned in which *E* and *S* could be alone. The apparatus, with all sections covered by the doors, was placed on a table between *E* and *S*. The *E* began by saying, "We are going to play a game. If you listen and do as I tell you, you will have a good time." The first phase of the procedure consisted of training on each of the three habit segments in one of the six prearranged orders. The directions for each segment were as follows. (The treatment differed only in the order in which they were presented.) For the A-B and X-Y segments *E* raised the door on either the right or left panel and said, "Here are two different handles. This handle moves like this, and this handle moves like this." The *E* demonstrated as she spoke and then gave *S* an opportunity to try the manipulanda. When *S* had mastered the manipulation *E* said, "Moving one of these handles will drop something in here. You move one of them and see if it is the right one, but remember, each time you may move only one. Then I will get the game ready like this for another turn." While speaking *E* demonstrated by replacing the door. A noncorrection procedure was used throughout, i.e., after each response the door was replaced, while *E* made or pretended to make the necessary adjustments. The time between trials was not held constant, but it tended to be of very short duration, somewhere between 5 and 15 sec. When *S* made a correct response, the subgoal (brass cube or glass marble)

appeared in the chute and *E* said, "You got it this time. Put it in this cup." The *E* pointed to one of two paper cups provided to hold each of the subgoals. "You won't be able to keep it, because I need it to play the game with other children. But you try to get one on *every* turn anyhow." *S* continued to respond to one panel until a criterion of three successive correct responses was reached.

The instructions for the B-G experience were as follows: After the screen on the central panel was removed *E* said, while pointing to the relevant parts of the apparatus, "See, here is some candy. Would you like to have it? Now look at this hole. See it? Now look at the things in the cups. Drop one of them in the hole and see if the candy comes out here. Drop only one in at a time. Then I will close the door like this and get the game ready for another turn. You try to get a candy on every turn." If *S* used the correct subgoal he got an M & M candy, which he was allowed to eat or to save if he preferred. On this panel, as on the others, responding was continued until a criterion of three successive correct responses was reached.

During training only one panel was opened at a time. When the criterion on any given segment was reached, that panel was closed. The *E* shifted *S*'s attention to the next panel by saying, "Now look at this part of the game."

After *S* had been trained on each of the three segments, the test situation was presented in the same manner to all *S*s. All panels were covered and *E* said, "Would you like some more candy? Well, I will empty these cups," referring to the subgoal containers, "but I will open the whole game." All doors were removed. "If you do the right things, you will be able to get some more candy. You try to get the candy." For the next 30 sec. *S* was free to make any response he chose, including no response. All responses were recorded.

In order to minimize contamination three factors were counterbalanced in the following manner. In each experimental group: (*a*) The right panel served as the A-B segment for half the *S*s, while the left panel was the X-Y segment. The opposite arrangement applied to the remaining *S*s; (*b*) Each of the two manipulanda of the A-B panel was correct for half of the *S*s; (*c*) For half of the *S*s the marble was B and the cube Y; for the remainder the condition was reversed.

Results and Discussion

The present apparatus, which is more complex than the one previously used, requires a somewhat new classification of *S*'s behavior during the

TABLE 1

Frequency Distribution of Initial and Integration Responses as a Function of Order of Presentation of Behavior Segments

Response Category	Order of Presentation						Total $N = 192$
	I $N = 32$	II $N = 32$	III $N = 32$	IV $N = 32$	V $N = 32$	VI $N = 32$	
	B-G A-B X-Y	X-Y B-G A-B	B-G X-Y A-B	A-B B-G X-Y	A-B X-Y B-G	X-Y A-B B-G	
Initial Responses							
Correct	8	11	14	12	11	13	69
Incorrect	12	15	9	10	13	11	70
None	12	6	9	10	8	8	53
Integration Responses							
Correct	2	4	6	6	1	6	25
Correct with unnecessary Rs	5	9	4	3	10	11	42
Incorrect or none	25	19	22	23	21	15	125

test trial. The test behavior is analyzed into two separate components: (*a*) *S*'s initial reaction to the test situation, and (*b*) *S*'s reactions after his initial choice. The scoring categories of each component will be described, including the number of *S*s responding in that manner.

Ss' initial reactions.—Three mutually exclusive categories of *S*'s initial reactions to the test were used (see Table 1). (1) *S* could respond correctly by moving the appropriate manipulandum on the A panel, thus securing B. Sixty-nine of the 192 *S*s responded this way. (2) *S* could respond incorrectly in one of three ways: (*a*) by choosing the incorrect manipulandum on the A panel; (*b*) by choosing the correct manipulandum on the X panel (i.e., the one that produced the Y subgoal), and (*c*) by choosing the incorrect manipulandum of the X panel. Seventy *S*s responded incorrectly with 6, 59, and 5 *S*s falling respectively in the *a*, *b*, and *c* subcategories. (3) *S* could react to the test situation by making no response to the apparatus. Fifty-three *S*s responded in this manner.

Ss' subsequent reactions.—The tendency to integrate the A-B and B-G segments to attain G was independent of the initial reaction. The *S* could make the initial correct choice and yet not complete the B-G segment. On the other hand, *S* could make an initially incorrect choice and then go on, in the time allotted, to respond correctly to A and finally complete the A-B, B-G sequence to get his candy.

Again three mutually exclusive categories were employed (see Table 1). (*a*) *S* could make the initial correct choice and, with no intervening responses to the apparatus, use B to get G. Twenty-five children displayed such smoothly coordinated inferential behavior. (*b*) *S* could complete the A-B, B-G sequence but make unnecessary responses along the way. Forty-two children responded in such a way. Of these 19 made only one unnecessary response, such as choosing X once or repeating the A-B sequence once before getting G. (*c*) The third category includes the 53 *S*s who made no initial response, 67 other *S*s who made an initial response but never reached the point of dropping B into G, and the 5 *S*s who dropped Y into G.

What these overall results show is that the problem was difficult for children under 6 yr., since most of them did not solve it. Nevertheless, there were some solutions. Thirteen percent of the *S*s solved it with maximum efficiency and an additional 22% reached solution albeit with unnecessary detours.

The effect of each order of presentation on the initial and subsequent components of the inferential sequence is shown in Table 1. The differences attributable to order on either of the components are not statistically reliable, according to the χ^2 tests of independence applied to to each table.

The fact that there are no reliable differences between any of the orders, when each one is taken by itself, may be attributable to the small *N*. One way to provide more stable tests of the influence of several independent variables is to enlarge the *N* by combining the relevant experimental groups. Table 2 presents in percentage terms the results of combinations dictated by considerations to be described. The following analysis will correspond to successive headings of Table 2.

Precedence of A-B or B-G.—The first hypothesis tested was the one generated from both the Hullian and Skinnerian

TABLE 2

PERCENTAGE OF INFERENTIAL RESPONSES ON TEST TRIAL ACCORDING TO FOUR DIFFERENT ORDER OF ACQUISITION VARIABLES

Order	Initial Response			Integration Response		
	Correct	Incorrect	None	Correct	Correct With Unnecessary Responses	Incorrect or None
Precedence of A-B or B-G						
B-G precedes A-B	34.4	37.5	28.1	12.5	18.8	68.7
A-B precedes B-G	37.5	35.4	27.1	13.5	25.0	61.5
Contiguity of A-B and B-G						
B-G contiguous with A-B	34.4	37.5	28.1	14.1	21.9	64.0
B-G and A-B separated by X-Y	39.1	34.4	26.6	10.9	21.9	67.2
Position of B-G						
B-G first	34.4	32.8	32.8	12.5	14.1	73.4
B-G second	35.9	39.1	25.0	15.6	18.8	65.6
B-G last	37.5	37.5	25.0	10.9	32.8	56.2
Terminal Subgoal Segment						
A-B terminal	39.1	37.5	23.4	15.6	20.3	64.1
X-Y terminal	31.2	34.4	34.4	12.5	12.5	75.0

type of analysis, namely, that the probability of inference would be greater when the B-G segment preceded the A-B segment than vice versa. The data for Groups I, II, and III were combined and compared with the combined Groups IV, V, and VI. Rather than emphasizing any differences, this combination produced results that are remarkable for their similarity. The very slight differences that do occur are in a direction opposite to the prediction.

Contiguity of A-B and B-G.—Underwood has stressed the importance of contiguity for problem solving. He writes, "The basic assumption is that in order for relationships to be perceived and acquired, responses to those stimuli must be contiguous" (Underwood, 1952, p. 212). Roughly translated to the present situation, this assumption would predict that inferential behavior would be greater when the A-B and B-G segments are contiguous. To test this possibility, the four experimental groups in which these segments were contiguous (I, II, IV, and VI) were combined and compared with the two remaining groups in which the X-Y segment was interposed between A-B and B-G. The differences obtained are such that the initial response results are in a direction opposed to the prediction, but the correct integration responses are in accord with it. The differences, however, are very small and the null hypothesis is highly tenable. Contiguity, as defined in this experiment, does not seem to be an important variable in the inferential behavior of children.

Position of B-G.—In an earlier study of inference (Kendler & Kendler, 1956), it was possible to make limited tests of the effect of the order of presentation variable. It was found that there were more initial correct (A) choices when

B-G was last than when it was first. (In that experiment there were no groups in which B-G was presented second, and there was no separate measure of the second component of the inference response.) To determine the generality of the observed relationship, the present data were compared for groups in which B-G was first (I and III), second (II and IV), and last (V and VI). Although the differences were not reliable, they are consistent with previous findings. There is a very slight increase in correct initial responses as B-G goes from first to last. There is a larger increase in the category of "correct integration responses with unnecessary steps." None of these differences achieved statistical significance.

Last subgoal segment.—In the initial study (Kendler & Kendler, 1956) a tendency was observed for *S* to alternate his first test response with his last training response. If, for example, the last training trial was X-Y, the probability of a correct initial test response (A) increased. Conversely, if the last training trial were A-B, the probability of an incorrect initial test response (X) increased. There was some suggestion that this effect was related to the specific apparatus and procedure employed, rather than to the inference process as such. This suggestion was supported in that in the present experiment with an improved apparatus, the tendency "to alternate" disappeared.

On the whole, the present results fail to lend any support to the idea that the order-of-presentation variable is an important one in the inferential behavior of young children. The results are also inconsistent with the extensions of the formulations of Hull (1935, 1952), Skinner (1938), and Underwood (1952) as they apply to the present experiment. These inconsistencies lend themselves to three general interpretations. The first is that the theoretical notions involved are inapplicable to the kind of problem-solving behavior that is being studied. The second possibility is that the fault lies not with the theoretical assumptions but the manner in which they were tested. Perhaps experimental operations that allowed a better control of *S*'s mediated responses would have produced results more in line with the predictions. The third possibility is that additional assumptions must be added to those tested to account for inferential behavior. Assuming bidirectional verbal associations would be one such possibility. It remains for future research and theory to decide which of these three alternatives is most appropriate.

Summary

An experiment was conducted on children between 30 and 65 mo. of age to determine whether inferential behavior resulting from the linkage of two discretely acquired behavior segments is influenced by the order in which these segments are presented during training. Explanation of inference by the invocation of either the anticipatory goal mechanism or secondary reinforcement would imply that order would be a variable. The results failed to confirm this implication. Since no statistically reliable differences were found to be attributable to order, the suggestion is that it does not have an important influence on inferential behavior in children.

References

Deutsch, J. A. The inadequacy of Hullian derivation of reasoning and latent learning. *Psychol. Rev.*, 1956, **63**, 389–399.

Hull, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, **42**, 219–245.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Kendler, H. H., & Kendler, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.

Kendler, H. H., Kendler, T. S., Pliskoff, S. S., & D'Amato, M. F. Inferential behavior in children: I. The influence of reinforcement and incentive motivation. *J. exp. Psychol.*, 1958, **55**, 207–212.

Maier, N. R. F. Reasoning in white rats. *Comp. psychol. Monogr.*, 1929, **6**, No. 29.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Underwood, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–219.

(Received April 14, 1960)